Essays on the Language-Cognition-Perception Interface

By
Hunter Gentry

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Philosophy)

At the
UNIVERSITY OF WISCONSIN-MADISON
2024

Date of Final Oral Examination: May 31st, 2024

The dissertation is approved by the following members of the Final Oral Committee:
Hayley Clatterbuck, Retired Faculty, Philosophy
Larry Shapiro, Berent Enç Professor of Philosophy, Philosophy
Farid Masrour, Professor, Philosophy
Gary Lupyan, Professor, Psychology
Bruno Whittle, Associate Professor, Philosophy

TABLE OF CONTENTS

DEDICATION

This dissertation is dedicated to my mother who has never stopped believing in me even when I had a hard time believing in myself. Mom, you inspire me in ways I cannot express. Thank you for being you.

ACKNOWLEDGMENTS

There are many people I would like to thank for supporting me during my graduate career. First and foremost, I am indebted to my advisor, Hayley Clatterbuck, who was a consistent source of support and inspiration throughout my career. Hayley has a remarkable way of clarifying and organizing my (sometimes) incoherent ramblings. I also want to thank my committee including: Larry Shapiro, Farid Masrour, Gary Lupyan, and Bruno Whittle. Thank you for agreeing to serve on my committee, supporting, challenging, and believing in me.

It would be a bad strategy to not thank my Houston Boyz: Jacob MacDavid, Mica Rapstine, Sam Ridge, and Jon Weid. Thank you for being such good friends. Thank you for listening, challenging, and loving me. Boyz night is a constant source of inspiration, but more importantly, a reminder of how lucky I am to have you all in my life.

Huge thank you to Stephanie Hoffmann too. Your endless support, respect, and care amaze me. I am so grateful to be part of your life.

Cameron Buckner's support, not only through my graduate career at UW-Madison, but also while I was a master's student at the University of Houston, was foundational for this dissertation. Thank you, Cameron, for always challenging me. You have an incredible mind, and I am grateful to have you in my corner. Thank you for all the guidance and opportunities you have afforded me.

Thank you to my best friend, and brother, William Schloetel. We discovered philosophy together. Our conversations have been formative in keeping my wonder and curiosity alive. You continue to inspire and challenge me. I am so grateful for our friendship.

I also want to thank all my fellow graduate students in the Philosophy Department at UW-Madison who have shaped my thinking in countless ways. Special thank you to Lukas

Myers and Rebecca Evans, Casey Ruefner, Madi Hass, Dylan Beschoner and Agustin Olavarria, Emma Prendergast, Katie Deaven, Hubert Marciniec, Greg Nirshberg, Jonathon VandenHombergh, Michael Bruckner, Elizabeth Bell, and Arun Baxter. Huge thank you to my office mates, Alec Michaels and Henry Curcio, for listening to and supporting me.

Thank you to all the UW-Madison Philosophy Faculty. Special thank you to Russ Shafer-Landau for all the help and support in applying for jobs. You kept me grounded and hopeful in a very tumultuous time. Special thank you to Jimmy Goodrich for the grounding conversations and advice.

Thank you to the Philosophy Faculty at the University of Houston as well. I am especially appreciative of the support and guidance from Justin Coates, Tamler Sommers, David Phillips, James Garson, and Josh Weisberg.

I want to thank all the Philosophy Faculty at the College of William & Mary. I am especially thankful for the support and guidance from Chris Tucker, Paul Davies, Aaron Griffith, Matthew Haug, Noah Lemos, and Elizabeth Radcliffe.

Thank you to Rick Alley, my English professor at Tidewater Community College. Thank you so much for being a friend and mentor. I can't express how much it means to me that you listened to and supported me.

Finally, I want to especially thank Michael Tarpey, my first philosophy professor at Tidewater Community College. You have no idea how much our relationship meant to me. Without you, I wouldn't have discovered my passion for philosophy. Thank you.

ABSTRACTS

CHAPTER 1: Constructing Embodied Emotion with Language: Moebius Syndrome and Face-Based Emotion Recognition Revisited

Some embodied theories of concepts state that concepts are represented in a sensorimotor manner, typically via simulation in sensorimotor cortices. Fred Adams (2010) has advanced an empirical argument against embodied concepts reasoning as follows. If concepts are embodied, then patients with certain sensorimotor impairments should perform worse on categorization tasks involving those concepts. Adams cites a study with Moebius Syndrome patients that shows typical categorization performance in face-based emotion recognition. Adams concludes that their typical performance shows that embodiment is false. Moebius patients must draw on amodal (non-embodied) emotion concepts. In this paper, I review face-based emotion recognition studies with Moebius patients yielding conflicting results, and diagnose these conflicts as a difference in experimental design. When emotion labels are provided, patients have typical performance, but when labels are not provided patients are severely deficient. I then show how an embodied, psychological constructionist view of emotions predicts and explains these performance differences. The upshot is that embodied theories of concepts are vindicated.

CHAPTER 2: What's in a Color?: Labels, Gesture, and Synesthesia

The label superiority effect, wherein subjects provided labels in discrimination and categorization tasks outperform those subjects who do not have labels, is a well known finding in cognitive psychology. A commonly accepted explanation for this effect is that language, and in particular, category labels, make perceptual representations more categorical. Perception is more categorical when between-category differences are more salient than within-category differences, as when the boundary between red and yellow is more salient than reddish-orange and reddish-yellow. Categorical perception involves the abstraction, or removal, of category-irrelevant features. Is language special in its ability to augment perceptual representations through abstraction? In this paper, I argue that it is not. At first, analog, bodily gestures seem to be good candidates of non-linguistic representational abstraction. This is because many

studies have shown that gestures promote learning and cognizing in various tasks. A commonly accepted explanation is that gestures "schematize" representations by abstracting away from task-irrelevant features. I argue that the jury is still out as to whether gesture qualifies as a form of *non-linguistic* representational abstraction. However, I then argue that olfactory-color synesthesia does qualify. Hence, language is not special in its ability to augment representations through abstraction. I end by providing evidence of a common mechanism subserving representational abstraction across language, gesture, and synesthesia.

CHAPTER 3: The Constructive Episodic-Semantic Memory System

What is the nature of semantic memory? Philosophers and cognitive scientists have long held that semantic memory stores invariant knowledge structures to be retrieved as such. In this paper, I argue that this conception of semantic memory is likely false. In particular, I argue that the neural mechanisms that realize episodic and semantic representations grade into one another, exhibiting the metaphysical phenomena of transitional gradation. I then motivate the consensus view regarding the constructive nature of episodic memory. That view says that episodic memory stores "traces", as opposed to the entirety of an event. At retrieval, the trace recruits various sensorimotor, affective, and spatial representations to reconstruct the target event as faithfully as possible. If it's true that episodic and semantic memory systems grade into one another, and it's true that episodic memory is constructive, then we have good reasons to think that semantic memory is constructive too. I close the paper by providing independent evidence that semantic memory is constructive, and providing a proof of concept for the view on offer– the semantic pointer architecture.

CHAPTER 1: Constructing Embodied Emotion with Language: Moebius Syndrome and Face-Based Emotion Recognition Revisited[1]

1. Introduction

One hypothesis of embodied cognition states that cognition centrally involves sensorimotor processes (Newen, de Bruin, and Gallagher, 2018).[2] This hypothesis has received widespread attention in recent philosophical and psychological discourse. Many psychological studies with surprising results have been taken to support embodied cognition (for a comprehensive review and discussion see Farina 2021; Gallagher 2011; Shapiro 2019). Of special interest to this strand of embodied research is the suggestion that concepts are embodied.[3] Concepts are embodied, on this view, in the sense that deploying them involves sensorimotor simulations which activate sensorimotor cortex.[4]

Consider the following case from Bergen and Feldman (2008):

---

[1] At the time of writing this, this chapter is forthcoming at the *Australasian Journal of Philosophy.*

[2] Following Shapiro (2019), there is probably not one thesis that unifies embodied cognition research. Indeed, there are at least three logically independent theses we can distinguish:
Conceptualization: the properties of an organism's body limit or constrain the concepts that that organism can acquire.
Replacement: Cognitive processes are not discrete, but continuous which makes the standard computational framework ill-suited to explain cognition. Additionally, an organism's body in interaction with the environment replaces the need for representational processes.
Constitution: The body or the world plays a constitutive rather than a merely causal role in cognitive processing (pp.4-5).
Replacement is most closely associated with anti-representational and dynamical systems approaches to cognition (e.g., Chemero 2011), whereas *Constitution* is most closely associated with extended cognition (e.g., Clark and Chalmers 1998). In this paper, I am focusing on *conceptualization*, and setting aside the latter two theses.

[3] For a discussion of learning embodied concepts see Bergan and Feldman (2008). See also Núñez (2008) for a discussion of embodied mathematical concepts.

[4] Embodied concepts are sometimes referred to as "modal concepts" to contrast with "amodal concepts". They are called "modal" because the vehicle of representation is grounded in sensory modalities. Amodal concepts are abstract, symbolic representations, e.g., Language of Thought (Fodor 1975). However, see Michel (2021) for skepticism regarding the distinction between modal and amodal concepts.

> Can you say how many windows there are in your current living quarters? Almost everyone simulates a walk-through to count them. Or consider a novel question—could you make a jack-o-lantern out of a grapefruit? (pg. 315).

According to Bergen and Feldman, in deploying your concept of a grapefruit, you engage in a task-dependent kind of sensorimotor simulation.[5] In this case, "reflecting on the carvability of a grapefruit involves creating internal motor and sensory experiences of carving a jack-o-lantern out of a grapefruit" (pg. 315). For these theorists (and many others), "Any time we use concepts, whether in performing categorization tasks, processing language about concepts, or reflecting on their features, we use mental simulation—the internal creation or recreation of perceptual, motor, and affective experiences" (pg. 315). The internal (re)creation of these experiences is thought to be implemented by activation of sensorimotor cortex.

Or consider the action-sentence compatibility effect (see Bergen 2012 for extended discussion). When asked to judge whether a sentence made sense, and to indicate their judgment via an action (here: pushing a button), subjects were faster at responding when the action was consistent with the target sentence. The hypothesized explanation is that interpreting the sentence requires sensorimotor activation, and concurrent inconsistent motor behavior interferes with this process. If interpreting sentences requires the deployment of concepts and interpreting sentences requires sensorimotor activation, then plausibly, what it is to deploy a concept is to activate sensorimotor cortex, i.e., embodied simulation.

Fred Adams (2010) has advanced an empirical argument against embodied concepts. Adams reasons that if concepts are embodied, and deployment of an embodied concept involves

[5] There is a wealth of literature on embodied simulation (Barsalou 1999; Bergen 2012; Shapiro 2019). For work on embodied simulation and metaphor see Gibbs, Jr. (2006); Lakoff and Johnson (2008).

simulating sensorimotor processes, then patients with certain sensorimotor impairments should perform worse on categorization tasks involving those concepts. Moebius syndrome patients suffer from congenital bilateral facial palsy due to cranial nerve underdevelopment. As such, these patients cannot make facial expressions. The embodied concepts thesis predicts that these patients would have impairments in face-based emotional recognition tasks because they would not be able to simulate facial expressions. However, in Calder et al.'s (2000) study, Moebius patients did not significantly differ in their performance on such tasks compared to healthy controls. Adams concludes that their comparable performance shows that embodiment is false. Moebius patients must draw on amodal (non-embodied) emotion concepts.

In this paper, I will attempt to show a few things. For one, I want to complicate the picture that Adams (2010) has about Moebius syndrome performance in face-based emotion recognition tasks. The experimental literature has some studies showing comparable performance to healthy controls and others showing deficient performance. I think these mixed results are the reflection of a key experimental design difference--the use of labels. When experimenters provide labels, Moebius patients perform comparably to non-Moebius subjects, but when labels are not used, the patients are unable to complete the task. I will explain these results via research on the language-cognition-perception interface that shows a "label superiority effect" among subjects in categorization tasks. While the importance of labels seems to lend support to Adams' view that Moebius patients are utilizing amodal concepts, I will argue that it is also compatible with a more embodied view of emotions-- psychological constructionism. I will describe the basic philosophical commitments associated with this view, and then show how the view predicts and explains the performance differences in Moebius

patients. The result is a view of embodied emotional concepts held together by the "glue" that is language.

Here's a roadmap for the rest of the paper. In section 2, I review a few different Moebius syndrome studies of face-based emotion recognition, paying special attention to the performance differences when labels are used or not used. I will then provide a brief analysis of the findings. In section 3, I will argue that the use of labels is the key difference in experimental design by reference to the "label superiority effect". In section 4, I will get specific about what role labels are playing in the Moebius syndrome studies such that Moebius patients can perform comparably to controls. Here, I will also introduce my favored theory of emotions: psychological constructionism and show how this view predicts and explains the performance differences.

2. Face-Based Emotion Recognition in Moebius Patients

Adams' (2010) argument against embodied concepts can be formalized as follows:

P1. If Moebius patients succeed on face-based emotion recognition tasks, then concepts are not embodied.

P2. Moebius patients do succeed on such tasks.

C. So, concepts are not embodied

Premise 1, Adams thinks, is what embodied theorists are committed to saying about Moebius patients. In particular, if the embodied theorist thinks that simulation of facial expressions is necessary for successful categorization of facial expressions, then because Moebius patients cannot simulate facial expressions, they will not succeed. Premise 2, Adams thinks, is supported by Calder et al. 's (2000) study showing successful categorization in Moebius patients. And I suppose he thinks that the results must generalize. The conclusion, then, follows deductively.

I take issue with both premises of this argument. I will start with premise 2 by reviewing a few faced-based emotion recognition studies with Moebius patients that yield conflicting results. I will ultimately argue that the performance differences are due to differences in experimental design-- the use of emotion labels. This difference leads to two different kinds of studies-- both probative, but eliciting different capacities. Importantly, these differences in performance can be predicted and explained by my favored view of emotions: psychological constructionism; therefore, premise 1 is false. But before I get to all of that, let us take a closer look at the complicated picture of Moebius patients' performance on face-based emotion recognition tasks.

2.1 A Complicated Picture of Moebius Patient Performance

Originally described by von Graefe and Saemisch (1880) and named by Moebius in 1888, Moebius syndrome is congenital bilateral facial palsy (and/or bilateral abducens palsy) resulting from cranial nerve underdevelopment (more specifically cranial nerves VII and VI).[6] Incidence of the syndrome is complicated due to comorbidities and inconsistencies of associated symptoms and features (Bell et al. 2019); however it is estimated to occur in 1 in 50,000 to 1 in 500,000 live births (Rasmussen et al. 2015). Some researchers report that there is equal incidence among genders (De Stefani et al. 2019; Nicolini et al. 2019), but Bell et al. (2019) report slightly higher incidence in males. Additionally, most patients are of "normal" (i.e., typical) intelligence (Nicolini et al. 2019) and cognitive development (De Stefani et al. 2019).

[6] Perhaps a better name is "Moebius sequence", however, because to date, there are no strict diagnostic criteria due to inconsistencies of associated symptoms (for a review see Bell et al. 2019; De Stefani et al. 2019).

Due to facial palsy, patients cannot make facial expressions. As such, patients often report difficulties in building social relations. A deficit in emotional processing is thought to explain this difficulty. In particular, patients are thought to be deficient in attributing emotional categories to facial expressions.[7]

To test this claim, Giannini et al. (1984) had a single Moebius patient watch videotapes of people at slot machines playing for one cent, ten cent, or twenty-five cent jackpots. The videotapes recorded the facial expressions at selection of the jackpot and the payout (or lack thereof). After watching, the patient was asked to determine how much money was at risk. To do this, the patient was instructed to press the button corresponding to their answer (either 1¢, 10¢, or 25¢). Because the video only showed facial expressions, the judgment was made based purely on facial cues.

Despite having a typical IQ and master's level education, "the patient could not respond at all to the task" (Giannini et al. 1984; pg.174). The patient described the experience as like trying to understand a foreign language. Approximately 300 controls had performed the task without any issues. The authors hypothesize that because the controls did not suffer from any neurological conditions, the patient's inability to complete the task is due to her Moebius syndrome. That is, the fact that the patient cannot make facial expressions is supposed to explain her inability to recognize them.

More recently, Nicolini et al. (2019) were primarily interested in testing autonomic responses in Moebius patients, but to get a baseline for emotion recognition, they showed video

[7] Caution must be urged in interpreting the studies reviewed here due to extremely small sample sizes. It is difficult to draw any definitive conclusions. This cuts both ways-- against my own analysis and argument, as well as Adams (2010). However, I am attempting to provide a more holistic analysis of the available data on Moebius syndrome and face-based emotion recognition than Adams.

clips of cartoon characters acting out emotions to be identified by the subjects.[8] Initially, they asked the patients what emotional state they thought the character in the video was in. Importantly, this was a free-response question with no cueing or labels provided. Similar to the patient in Giannini et al.'s study, the patients could not respond to the questioning.

To overcome this difficulty and achieve their baseline measure, Nicolini et al. administered the Test of Emotion Comprehension (TEC) (Pons et al. 2002). The TEC typically involves the experimenter reading a story to the subjects or showing a video and then presenting the subjects with four facial expressions, e.g., sad, happy, angry, neutral. Nicolini et al. presented the four facial expressions and asked the patients to select which one matched the cartoon character's emotional state in the video they watched. Again, crucially, the presented facial expressions did *not* have emotion labels. The patients' performance on TEC was significantly impaired compared to controls (mean control scores: ~5 versus mean patient scores: ~2.5).

However, in 2000, Calder et al. conducted a face-based emotion recognition study with Moebius patients that showed comparable performance to healthy controls. Calder et al. (2000) tested three Moebius patients on a battery of tasks. Most important for our purposes, however, is the Ekman facial recognition task. The researchers used the Ekman and Friesen (1976) series of 60 photos. There were 10 models who each assumed 6 different facial affectations (happiness, sadness, anger, fear, disgust, and surprise). The subjects were asked to identify which emotion was being modeled in each picture from a list of 6 emotion labels. Patient performance did not

[8] Nicolini et al. (2019) say that the "Stimuli were comprised of short clips taken from the Internet in which the main character of the scene was in a happy, sad, or scary situation" (pg. 3). Apparently, the video clips included more than just facial expressions which is significant because the subjects could not respond to the questioning. This might be a reflection of children's underdeveloped emotional conceptual system in general, which could be severely underdeveloped in Moebius children. See Widen (2013) for a discussion and review of children's emotional conceptual development.

significantly differ from 40 healthy controls (mean patient score: 47.67/60.00 versus mean control score: 50.35/60.00). However, Calder et al. do note that patients' performance was slightly deficient.

Bogart and Matsumoto (2010) also found that Moebius patients did not significantly differ from controls in face-based emotion recognition. At the time of my writing this, these authors have the title of largest sample size with Moebius patients– clocking in at n=37. Bogart and Matsumoto used a variant on the traditional Ekman set of photos, called the "Multi-Ethnic Facial Expression Set" (Matsumoto and Ekman 2006).[9] Procedurally, participants are presented with a photo of a face expressing one of seven emotions. They have unlimited time to view the photo and then choose the correct emotion label. The control group had a performance range of 52-93 (M=75.44) while the Moebius group ranged from 60-90 (M=73.67).

2.2 Taking Stock

So here's the complicated picture of Moebius patient performance on face-based emotion recognition tasks: Calder et al. (2000) and Bogart and Matsumoto (2010) report comparable performance to healthy controls (as do Vannuscorpus et al. 2020), but Giannini et al. (1984) and Nicolini et al. (2019) report deficient performance.[10] These mixed results, whatever the explanation, cast doubt on Adams' 2nd premise. It is not the case that Moebius patients always succeed on emotion recognition tasks. They sometimes do; and in the studies where they do succeed, the experimental design is different from the studies where they do not. In particular,

---

[9] The original Ekman task used only white models posing in 6 basic emotions. This variant used black, brown, and white models posing in 7 basic emotions, adding "contempt".

[10] Giannini et al. did not make use of *emotional* labels. The patient had to recognize the emotional expressions in the videos and then infer how much money was at stake. It is doubtful that labels for the amount of money would interact with emotional processing to produce an advantage, unless the patient knew the meaning of the emotional expressions already.

Calder et al. (2000), Bogart and Matsumoto (2010), and Vannuscorpus et al. (2020) provided a list of emotion labels to choose from, whereas Giannini et al. (1984) and Nicolini et al. (2019) did not use emotion labels-- subjects free-named the facial expressions.

In addition, Nicolini et al. (2019) tested young children with Moebius syndrome as opposed to adults. Here's what they say about this decision:

> Our study is the first to use a relatively large sample of very young patients to investigate the effects of facial muscle paralysis on both autonomic responses and emotion recognition. The investigation of these issues early in development *is critical for the detection of emotional processing mechanisms at a stage where more complex cognitive strategies might not yet compensate for their deficits* (2019, pg.8; emphasis added).[11]

The claim here seems to be that typical (as opposed to deficient) adult Moebius performance on emotional recognition tasks might be a reflection of compensatory mechanisms. It is possible that adult patients have developed strategies for emotional recognition to compensate for their inability to make and simulate facial expressions (Michael et al. 2015). So success on these tasks could have other explanations.

Correlatively, *deficits* on emotion recognition tasks could have other explanations as well. For example, the difficulties we see in social interaction between Moebius subjects and non-Moebius subjects could perhaps be due to the non-Moebius subjects having difficulty in the interaction. As Michael et al. (2015) put it, "Since people are accustomed to receiving information about others' mental states from their facial expressions, the absence of this

[11] De Stefani et al. (2019) found similar findings with children with Moebius syndrome. In a forced choice emotion categorization task (with stylized faces rather than labels), Moebius children were less accurate than controls. The authors hypothesize that the children might rely on more rule based strategies to identify the target emotion. For example, if the stimuli has feature F, then it is emotion E. The problem with such a strategy is that if F is shared with other emotional categories, then F is not diagnostic of a *single* category. As we will see in later sections, there is reason to suspect that features of emotional facial expressions are shared with many emotional categories.

expected information may interrupt an interaction partners' facial mimicry and cause him or her to feel uncomfortable or confused about what the person with [Moebius syndrome] is thinking or feeling" (pg.2). Furthermore, as an anonymous reviewer pointed out, Moebius subjects may have received less social input as children because others deemed them to be uninteresting or unresponsive.

This is all to say that success and failure on these face-based emotion recognition tasks are hard to interpret. However, this is a significant point because it puts pressure on Adams' premise 1. In particular, *adult* patients succeeded in the studies reported above, and this could be due to compensatory mechanisms that are themselves embodied. Without ruling this possibility out, Adams cannot conclude that embodied theories as a whole are mistaken. That is, Moebius patients can succeed on emotion recognition tasks and embodiment can still be true. So, given that when labels are used patients succeed, and when labels are not used they fail, could it be that the labels are functioning as a compensatory mechanism? Further, could it be that labels interact with embodied processes?

3. The Label Superiority Effect and the Language-Cognition Interface

In this section, I will begin to make the case that the use of labels in the Moebius syndrome studies is a key difference. Specifically, it could be the case that the use of labels explains how Moebius patients are able to successfully categorize emotional facial expressions. Moreover, as I will argue later, this explanation, although involving labels, is still embodied. I will start by discussing the importance of labels in categorization-- including emotional categorization-- in healthy populations. I will then review some evidence that emotional categories are perceptually fuzzy and complex, and that language helps to simplify distinctions between categories.

3.1 How and When Labels are Cognitively Advantageous

Carruthers (2002) distinguishes between strong and weak theses concerning the interface between cognition and language. Strong theses have it that language is somehow necessary for *all* human cognition. Weak theses have it that language is not necessary for cognizing, but that it "scaffolds" cognition–language is necessary to acquire some (but not all) concepts and/or language reduces computational burdens. I will not be taking a stance on whether the strong thesis is true, but I will be assuming that the weak thesis is true.[12] Everyone can agree that a human with language and another human without language will differ cognitively. The question is: how different will they be?

A perceptually fuzzy category is a category whose perceptual presentation overlaps with other categories. Importantly, this does not mean that the underlying reality of the category is indeterminate, it just means that evidence for the correct application of the corresponding concept is fuzzy or complex. Here's an example: COVID-19 and the common cold are distinct diseases, but they have overlapping symptoms. This makes it hard to perceptually distinguish the two, but this does not mean that there is no boundary between the two disease categories.

One function that labels play is to transform perceptually fuzzy or complex categories into perceptually discrete, unified categories. This is a function that language plays when learning the target concept as well as deploying the target concept after it has been learned. When a category is perceptually fuzzy, graded, or probabilistic, it might not be possible to learn the corresponding concept without language. Extreme examples might include 'democracy' and 'charity'– these highly abstract concepts do not have simple perceptual extensions. At the very

[12] Carruthers dismisses the strong thesis as it is formulated here.

least, it will be easier to learn these concepts with labels, than without.[13] Secondly, given that one has learned a perceptually fuzzy concept, having a label associated with it makes categorization easier. The label becomes associated with the most diagnostic features of that category, making salient boundaries between categories that are otherwise perceptually obscured, and serving as a direct route to category knowledge. In learning a word, one learns to chunk together certain, perhaps diverse, mental representations which facilitates concept learning and deployment (Pietroski 2005).

There are many views on how labels have this transformative effect (I am going to remain agnostic on this).[14] One view, call it "the attentional view", says that labels guide feature attention to the most diagnostic features of the category.[15] As an example, consider the shape bias. Starting around 18 months old, children will attend to shape as an indicator of category membership.[16] The idea here is that objects in the environment do not frequently change shape without also changing category. For example, I can't turn into a wolf without changing my

---

[13] Ramscar et al. (2010) argue that learning to predict a category label from exemplar features (FL learning) (rather than predicting features from a label (LF learning)) is more effective when the category is perceptually fuzzy. This is because, "LF learning tends to produce representations in which a number of competing outcomes are all highly probable" (pg.922). However, in FL learning, the subject discriminates between competing outcomes. The authors ultimately argue that language learning and understanding is a predictive process-- it's about tightly coupling the exemplar features that best predict the category label. In language understanding then, words are cues to predict meaning (see also Elman 2009).

[14] Another view says that language works to reduce the dimensionality of representations effectively making them more categorical. See Blouw et al. (2016; especially pg. 1138) , DiCarlo and Cox (2007), and Stewart et al. (2011) for explanations at the neural level. For views like this based on deep convolutional neural networks and transformer models see Buckner (2018), Gregor et al. (2016), and Valeriani et al. (2023). For a view like this based on psychological findings see Perry and Lupyan (2014). Finally, see Edelman and Intrator (1997a,b) for early descriptions of this view based on psychophysical findings from color and shape perception.

[15] Green (2020) explains the attentional view as a modulation effect: "A modulation effect makes a difference to the perceptual representation of feature values along a fixed dimension or set of dimensions (e.g., color, size, distance, contrast) that a perceptual process already has at its disposal" (pg.327).

[16] It is worth highlighting that it seems the shape-bias plays a role in Lupyan et al.'s (2007) alien study too (reviewed below). The fuzzy perceptual distinction between the two aliens consists in differences in shape.

shape. Children as early as 18-months old learn that despite perceptual dissimilarity along many dimensions between two stimuli, if shape stays constant, then they are likely of the same category.

In Linda Smith's "dax" experiments (Smith and Heise 1992; Jones and Smith 1993), young children are shown a "dax", and fail to extend the name to objects that do not share the same shape as the training exemplars. This suggests that shape matters to whether an item belongs to the dax category or not. However, they will extend the name if, for example, the size or texture changes, suggesting that these features are not important for membership.[17] Here, shape is not only useful in learning novel categories, it is also useful in deploying concepts after they have been learned too. If shape, in general, is a diagnostic feature of category membership across multiple categories, then attending to specific shapes will be useful in learning new concepts, as well as, deploying already learned concepts. Attending to shape, at the exclusion of other features of a particular stimulus, simplifies the representation of that stimulus. And when the attended-to feature is particularly diagnostic of a category, inferring category membership becomes easier (in learning or in deployment of learned categories).

In discussing these advantages, Clark (1998) says that "[T]he presentation of the same label accompanying a series of slightly different perceptual inputs (e.g., different views of dogs)...flags the presence of some further underlying structure and thus invites the network to seek the perceptual commonality" (pg.7). McClamrock (1995) makes a similar point when he says that labels thrust stable properties onto a complex environment effectively reducing our computational burden. In short, labels serve as a kind of "glue" to bind together perceptually

[17] They will not extend the name if the texture changes when the training sample has eyes. Smith and colleagues conclude that texture matters to extension of names if the exemplars are animate objects (the eyes are supposed to denote animate object).

complex stimuli, making them perceptually simple.[18] This makes concepts easier to learn while also facilitating categorization once the concept has been learned. This is one thing we mean when we say that "language scaffolds cognition".[19]

3.2 Experimental Evidence of the Label Superiority Effect

To help illustrate both of these advantages, let us consider the following study by Lupyan et al. (2007) which showed that subjects were faster and more accurate at categorizing novel stimuli when provided with a nonsense label than subjects not provided a label.[20] Lupyan et al. trained subjects to associate two kinds of "aliens" with either of two behavioral responses: "approach" or "flee". The two kinds of aliens were subtly different in shape (see picture 1). There were two training conditions: a label condition, in which the presented alien was labeled (either "leebish" or "grecious") and a no-label condition. Subjects were given feedback after each answer. After training, subjects were tested on how well they learned the associations. The results showed that subjects in the label condition, despite having the same amount of experience with the stimuli as subjects in the no-label condition, were faster and more accurate in both the training phase and at test. Lupyan et al. conclude that "learning nonsense verbal labels facilitated categorization more than did learning nonverbal associations" (2007, pg.1081). Lupyan et al.

---

[18] Some researchers have thought that labels "chunk" complex perceptual representations into discrete, unified representations (Huang and Awh 2018). This view might not be a competitor to the attentional view because attending to one perceptual feature could be an instance of chunking the perceptual representation. Alternatively, it could be that attending to category diagnostic features across perceptually dissimilar category exemplars allows the subject to chunk representations of those exemplars under a common label. See Pietroski (2005).

[19] It is worth noting that Clark is heavily influenced by Dennett (1984; 2015) and Deacon (1997). Early versions of this kind of view of language's influence on cognition can be found in these works. Thanks to an anonymous reviewer for this point.

[20] They were more accurate in the sense that there was a matter of fact about similarity that made a particular stimulus a member of one group and not the other. This is because Lupyan et al. created the stimuli to divide into two categories.

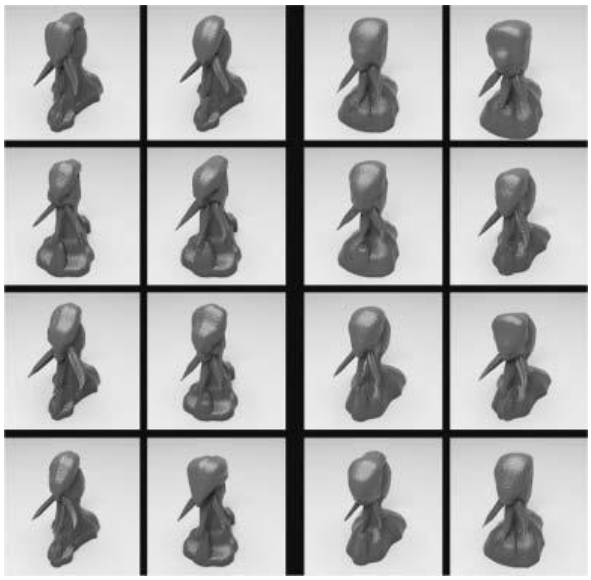
provide a possible explanation for the performance advantage: labels become associated with the most diagnostic information for that category. In their words:

> [R]ather than being fixed features, category names modulate item representations on-line through top-down feedback. According to this account, as a label is paired with individual exemplars, it becomes associated with features most reliably associated with the category. When activated, it then dynamically creates a more robust category attractor (pg.1082).

The idea here, echoing Clark (1998) and McClamrock (1995), is that labels facilitate learning and categorization by becoming strongly tied to the properties that distinguish that category from others.[21] Moreover, though the distinction between two categories might turn out to be perceptually complex, labels allow one to represent that distinction in a simpler way: "...labels may have allowed subjects to more easily represent the somewhat fuzzy perceptual distinction between categories (‘‘more rounded and smooth’’ vs. ‘‘less rounded, with ridges’’) in terms of a simpler verbal distinction (‘‘leebish’’ vs. ‘‘grecious’’)" (Lupyan et al. 2007, pg.1082).[22]

[21] See Casasanto and Lupyan (2011).

[22] See also Lupyan and Bergen (2016) and Lupyan and Thompson-Schill (2012).



PICTURE 1. The "aliens" on the left have flatter bases and a small ridge on the head. The "aliens" on the right have rounder bases and smoother heads. From Lupyan et al. (2007), pg. 1078.

3.3 Are Emotional Categories Discrete?

Dating back to Darwin (1872), emotions were thought to have a discrete, unified reality in nature. For example, Darwin thought that certain facial expressions and movements corresponded to specific emotional categories because they were adaptive (see Allport 1924). Inspired by Darwin, basic emotion theories (Ekman and Cordaro, 2011; Panksepp, 2011; Shariff and Tracy, 2011; for discussion see Deonna and Teroni, 2012, ch.2) state that emotions and emotional perception are innate, pre-linguistic, and universal.[23] Indeed, the inventor of the Ekman (1992) facial recognition task was a basic emotion theorist who thought that there were 6 basic emotions that corresponded one-to-one to certain facial expressions. It's not a coincidence

[23] There is ambiguity in the term "basic emotion". Scarantino and Griffths (2011) distinguish between three notions of "basic": conceptual, biological, and psychological. Something is conceptually basic just in case it occupies the basic level of a conceptual taxonomy. Something is biologically basic just in case it has an evolutionary origin and distinctive biological markers. Finally, something is psychologically basic just in case it does not contain another psychological component as a part. These notions of basic-ness are independent of one another. An emotion can be conceptually basic, but this does not entail that it is biologically or psychologically basic, and vice-versa. The basic emotion theorists, I think, have in mind biological basic-ness. It can be said that the Stoics and later early modern philosophers (e.g., Descartes, Spinoza, Hume, and Locke) thought that there were psychologically basic emotions (see Colombetti 2014, chapter 2).

that the Ekman facial recognition task has 6 different emotions represented by 6 distinct facial expressions (contempt has been added as a 7th basic emotion; see Matsumoto and Ekman 2004). If basic emotion theories are right, then emotions are *not* perceptually fuzzy categories. Rather, they are discrete, unified natural kinds. Further, if emotions correspond one-to-one to facial expressions, then, by hypothesis, seeing an emotional expression on a face should be sufficient information to know what emotional state that person is in.

Basic theories have faced challenges going back at least 50 years (Lacey 1967; for a review and discussion see Barrett 2006; Colombetti 2014). In what follows, I will review 2 main lines of experimental evidence that allegedly cast doubt on the basic emotion theorists' claim that emotions have a discrete, unified reality.[24] This evidence supports the claim that emotions are perceptually fuzzy categories. I will ultimately argue that because emotions are perceptually fuzzy categories, emotion labels are highly influential for everyone (healthy subjects and certain patient populations) in emotional recognition tasks. This should be sufficient to show that the use of emotion labels is a key difference in experimental design of the Moebius syndrome studies I discussed above.

To pump the reader's intuitions, I invite you to try the Ekman facial recognition task for yourself without labels (see picture 2). Take a moment and try to figure out what emotion is being expressed in each of the six photos.

---

[24] Above, I distinguished between a category being perceptually fuzzy and a category itself being fuzzy. I do not think the evidence canvassed here shows the latter unless one thinks that facial expressions, vocal signals, and various physiological responses are *constitutive* of emotional categories. If one thought that facial expressions, vocal signals, etc. were merely expressions of emotional categories, then the evidence canvassed here only shows that emotions are perceptually fuzzy. I will not be weighing in on the debate about whether emotional behaviors are constitutive or expressions of emotional categories. All I need to show for my argument is that emotions are perceptually fuzzy. However, if it turned out that emotional categories are *themselves* fuzzy, my argument would still go through.

PICTURE 2. Example photos from the Ekman Facial recognition task without labels.

From left to right starting at the top, we have anger, fear, disgust, shock, happiness, and sadness. How did you do? If you were barely above chance, you are not alone, and for good reasons. Barrett (2006) provides a review of emotion research that allegedly challenges the basic emotion theorists' claims that (i) emotions are discrete, unified categories (see footnote 23) and that (ii) emotion recognition proceeds automatically and passively. First, if basic theories are right, then there should be strong correlations between emotional categories and various measurable responses (e.g., subjective experience, facial expressions, and vocal signals). For example, when anger erupts in a person, there should be specific and distinct bodily responses: an increase in blood pressure, a scowl on the face, and perhaps the urge to yell or hit something. Unfortunately, these strong correlations, as predicted by the basic theories, have failed to materialize. Barrett concludes, "Taken together, enough evidence has accumulated for some theorists to conclude that lack of response coherence within each category of emotion is

empirically the rule rather than the exception" (2006, pp.33-34).[25] Essentially, these data suggest that there is variability between subjects in their bodily responses and subjective experiences of emotions, instead of consistent and distinct responses and experiences.

Another line of evidence allegedly against basic emotion theorists are production studies of emotion. In this kind of study, researchers measure bodily responses during emotionally triggering events to determine if there are signature responses for particular emotion categories. For example, researchers might be interested in whether there are signature facial muscle movements that underlie specific emotions. To do this, they try to elicit the target emotion and measure facial muscle movements using EMG. If basic emotion theories are right, then there should be distinct and consistent bodily responses that are produced for each emotion. Commenting on production based studies of the face, Barrett notes "that the bulk of existing evidence has failed to support the hypothesis of distinct patterns of automatic facial EMG activity for anger, sadness, fear, and other emotion categories" (2006, pg.39). Strikingly, this evidence is consistent with non-human animal studies that show that these creatures "rarely produce involuntary, reflexive displays" (Barrett 2006, pg.39).

On reflection, these findings are fairly intuitive. People do not always smile when they are happy, and people can be happy while furrowing their brow. A single facial expression can be associated with multiple emotional categories.[26] This observation has led Seyfarth and Cheney (2003) to conclude that because a single facial expression, in isolation, underdetermines what emotion the emoter is in, facial expressions (as well as vocal behaviors) have low

---

[25] Barrett cites (Bradley & Lang, 2000; Russell, 2003; Shweder, 1993, 1994) on this point.

[26] This does not mean that there is no connection between facial expressions and emotions. In general, there is a family resemblance inter- and intra-personally among facial expressions for particular emotional categories (see Barrett 2006, pp.37-39), but family resemblance is not what basic emotion theories predict.

informational value and low referential specificity. Importantly, the claim is not that facial expressions underdetermine what valenced state (i.e., negative or positive affect) the emoter is in. People are good at reading valence from facial expressions. They are not good at reading emotion, and that's because specific facial expressions do not correspond one-to-one to particular emotional categories. Taken together, this evidence shows why the reader (and the general population!) has difficulty on the Ekman facial recognition task (figure 2). Emotions are perceptually fuzzy and complex, and this makes emotional concepts especially good candidates for linguistic influence. In particular, labels allow everyone to represent perceptually fuzzy distinctions between emotional categories in a simpler way. By being associated with the most diagnostic features of emotional categories, labels become robust category attractors that bind together disparate perceptual information into a discrete and unified whole, while also serving as direct routes to category knowledge. I now turn to reviewing two main lines of evidence of linguistic influence on emotional processing.

3.4 Label Superiority Effect in Emotional Recognition Tasks

Russell and Widen (2002) provide good evidence that there is a label superiority effect in emotional recognition. Children ages 2 to 7 were asked to categorize various facial expressions under three conditions: label, face, or both. In each condition, a box was assigned a label, a face, or both and the children had to decide if the presented image went into the box or not. For example, if the box was labeled "happy", the child had to decide if the image of a smiling person went in the box. In the face only condition, the boxes were assigned pictures of facial expressions and the children had to decide which box another picture of a face went into. Children in the label-only condition outperformed those in the other two conditions. The authors conclude, similar to Lupyan et al. (2007), that emotional labels facilitate categorization.

Moreover, Russell and Widen comment: "These results invite further research on the assumption that facial expressions play a key role in the development of children's understanding of emotion. For instance, facial expressions might play an important role, *but only in combination with situational, vocal or behavioral cues*" (2002, pg.48). The idea here is that other sources of information, including linguistic information, very likely play an important role in emotional concept development. This last point is consistent with the evidence reviewed above against basic emotion theories.[27]

Another line of evidence comes from an observation from Lindquist and Gendron (2013):

> Although response options are typically considered an innocuous feature of the task, it has been shown that including emotion words in the experiment inflates participants' "accuracy" at identifying the emotion on the face…participants are generally better than chance at "accurately" identifying the emotion on the face when words are available in the experiment as response options (>63%). Studies that do not include emotion words in the task find substantially lower "accuracy" rates, however. For instance, the "accuracy" of responses is quite low when participants are asked to freely label an emotional caricature without being given a set of words to choose from (e.g., between 7.5% and 54%)... More strikingly, providing labels can even cause participants to perceive a face as an instance of an "incorrect" emotion (e.g., participants perceive a scowling face as "disgust" rather than "anger" when the word "disgust" is available but "anger" is not (pg.68).

Lindquist and Gendron are talking about *healthy* subjects' performance here. With labels, healthy subjects are above 63% accurate, but without labels healthy subjects' performance ranges from a staggeringly low 7.5% to 54%. Again, echoing Barrett (2006), face-based emotional recognition is really hard for everyone! Taken together, these two lines of evidence seem to weigh in favor of the hypothesis I floated above: given the data suggesting that emotional categories are perceptually fuzzy, one would expect emotion concepts to be highly

[27] It's also consistent with the larger theoretical point about labels–that they bind together complex perceptual stimuli making concept learning and later conceptual deployment easier.

susceptible to linguistic influence, e.g., label superiority effect. Language scaffolds emotional processing by transforming perceptually fuzzy boundaries between categories into perceptually discrete boundaries, and they do this by binding together perceptually disparate information into unified wholes.

3.5 Taking Stock

At this point, the reader is likely wondering how this discussion bears on the mixed results of the Moebius syndrome studies. Here's the upshot: healthy subjects are not great at face-based emotional recognition tasks even with labels, but much worse without. Moebius patients are worse still without labels. Given that both populations have access to language, the only difference between them that could explain Moebius patients' worse performance, when labels are *not* provided, is the fact that they cannot simulate facial expressions.[28]

However, even if Moebius patients are benefitting from a label superiority effect, and that's what explains their ability to succeed when labels are provided, it is not clear that that helps the embodied theorist. This is because labels seemingly activate amodal representations, not sensorimotor reenactments. After all, labels are themselves amodal representations-- they are symbolic, abstract stand-ins that arbitrarily pick out worldly things. So it's not clear how this shows that Adams' premise 1 is false. Indeed, one might think that this lends *prima facie* support to Adams' premise 1. If Moebius patients succeed on face-based emotion recognition tasks *via labels activating amodal representations*, then embodiment is false.

I think this move is too hasty. In the next section I will be laying out my favored embodied theory of emotions-- psychological constructionism-- which is compatible with the

[28] Although see the end of section 2.

label superiority effect, but rejects the idea that labels act to activate amodal representations. In other words, it could be true that Moebius patients succeed on face-based emotional recognition tasks using labels, but that does not show that embodiment is false. In fact, psychological constructionism predicts and explains Moebius patients' mixed performances.

4. Psychological Constructionism about Emotions

In this section, I will be introducing psychological constructionism in three parts. I will then show how this view explains the Moebius patients' mixed performance. I will begin with a brief (selective) survey on the philosophy of emotions.

4.1 Selective Survey of the Philosophy of Emotions

A natural objection to raise to Ekman-style basic emotion theory is that there are many more emotions than anger, fear, sadness, happiness, surprise, disgust, and contempt. What about love? Schadenfreude? Some basic emotion theorists distinguish between basic and non-basic emotions. There are, broadly speaking, two views on the relation between the basic and non-basic emotions: the unity thesis and the disunity thesis (Prinz 2004a; for review and discussion see Colombetti 2014). The unity thesis says that basic emotions are constitutive parts of non-basic emotions, and in virtue of this constitutive relation, basic and non-basic emotions are unified under the natural kind "emotions". The disunity thesis denies this– basic emotions are different in kind from non-basic emotions such that different theories must be given for each.

A representative disunity theorist would be Griffiths (2008) who thinks that basic emotions are homologues to certain structures in non-human animals. Non-basic emotions, on the other hand, are uniquely human. This fundamental distinction, for Griffths, means that basic

and non-basic emotions form two distinct natural kinds for which two different theories must be given.

A representative unity theorist would be Prinz (2004a) who thinks that all emotions are embodied appraisals. Embodied appraisals are perceptions of changes in the body that have the function of representing specific properties in the environment, e.g., danger in the case of fear, loss in the case of sadness. For Prinz, basic emotions are products of natural selection. That is, basic emotions are perceptions of bodily changes that have been naturally selected to have the function of representing a restricted set of properties. Non-basic emotions, on the other hand, are either "blends" of basic emotions or "recalibrated" basic emotions. By blending fear and surprise, for example, you might get the nonbasic emotion of awe. Prinz thinks that schadenfreude is a recalibrated basic emotion, namely joy. Joy has been recalibrated to represent others' suffering, instead of what it was naturally selected to represent.

Prinz's embodied view is inspired by William James (1884) and Carl Lange (1885). Both James and Lange thought that "emotions occur when the perception of an exciting fact causes a collection of bodily changes, and 'our feeling of the same changes as they occur IS the emotion'" (Prinz 2004b, pg.1). Prinz agrees that emotions are perceptions of bodily changes, but disagrees in that he thinks emotions are amenable to rational assessment. Hence, he thinks emotions are somatic, but also semantic–they represent contents as embodied appraisals.[29] The account I will offer later–psychological constructionism– is an embodied account that rejects basic emotions.

Prinz's view, especially the notion of "blending" basic emotions to make non-basic ones, is interesting because it could perhaps accommodate the data that constructionists levy against

[29] Slaby (2014) offers an account of extended emotion as an extension of the extended cognition hypothesis.

basic theories. The idea is that Prinz can agree that many emotions are perceptually fuzzy– the non-basic ones. The blending of basic emotions entails the overlap of emotional categories. Here, Prinz would agree that emotions are embodied, he would even agree that many emotions are perceptually fuzzy, but he would disagree that this means we should abandon basic emotion theory.

However, the perceptual fuzziness of emotional categories is not restricted to merely non-basic emotions. Constructionists think that perceptual fuzziness trickles down to even the supposedly basic emotions. In reaction to the findings canvassed in section 2.3, constructionists, "concluded that emotions are states tailor-made to a given context, which emerge when more elemental processes such as basic hedonic feelings or feelings of arousal are made meaningful using cognitive interpretation" (Lindquist 2013, pg. 357). Emotions are not basic, valence and affect are.

As evidence for this claim, consider the following. When researchers gather self-reports of subjective experiences of emotions and project them into a geometric space they do *not* find discrete clustering around the 6 putative emotional categories that would be indicative of distinct experiences among emotional categories. Rather, they find clustering around similarly valenced states, e.g., negatively valenced states like anger, fear, sadness. This suggests, comments Barrett (2006), that emotions are not experientially primitive (i.e., basic). That is, contra basic emotion theories, emotional experience is composed out of more basic psychological components, e.g., affect and arousal. Barrett goes further in her analysis:

> The main evidence that experiences of emotion can be broken down into more elemental bits is that when projected into geometric space, self-reports conform more or less to a circumplex structure (Feldman, 1995b; Remington, Fabrigar, & Visser, 2000; Russell, 1980; for a review, see Russell & Barrett, 1999). A circumplex structure emerges only

> when the objects in a correlation matrix (in this case, reports of emotion experience) are heterogeneous and psychologically reducible to a more basic set of properties (Guttman, 1957) (Barrett 2006, pg. 35).

There are two main takeaways from this line of evidence. The first is that if emotions are perceptually discrete categories with distinct and consistent behavioral and psychological responses, then subjective experience of emotions should reflect that. But self-reports do not conform to this categorical structure. Instead, and this is the second point, self-reports conform to more basic psychological components including affect and arousal. Self-reports cluster when projected into geometric space, just not in the way predicted by basic theories.

4.2 What is Psychological Constructionism? Part 1: Embodiment

In what follows, I will be discussing and motivating a particular kind of psychological constructionism about emotions that comes from Maria Gendron, Lisa Feldman Barrett, and Kristin Lindquist-- the conceptual act theory (or CAT). CAT is an embodied view that supplies a necessary role for language. I will argue that CAT provides a plausible explanation of the Moebius patients' performance. Hence, if true, the CAT account would constitute a counterexample to Adams' premise 1. In this sense, embodiment actually gains strong support from the Moebius syndrome studies.

If Barrett (2006) (and other constructionists) are right that emotions are composed out of more basic components, then how do emotions come about? CAT states that emotional experience and perception arise out of conceptual acts. A conceptual act, "is synonymous with categorization and relies on representations of prior experiences (i.e., knowledge, concepts, episodic memories). These prior experiences are represented as situation-specific reenactments of emotion in the brain's sensorimotor cortices" (Lindquist 2013, pg.362). Here, CAT is committed to the embodied approach to emotional concepts (that I outlined in the introduction).

That is, what people know about emotion is "represented in part by the same neural substrates that have increased activity when a person actually experiences core affective feelings, engages in behaviors, and perceives sensory stimuli" (Lindquist 2013, pg. 362). This means that when people have an emotional experience, say of anger, they are engaging in a conceptual act that is embodied. They categorize their experience as "anger" by sensorimotor simulation.[30]

The locus of the conceptual act need not be limited to the subject's emotional experience. One can also turn the locus of conceptualization outward onto the world as when one reads emotion on someone else's face. In both cases, "emotions emerge in consciousness when people categorize ambiguous internal and external sensations as instances of discrete emotion categories" (Lindquist 2013, pg.360). And in both cases, the conceptual act will be an embodied sensorimotor reenactment.

It's worth contrasting CAT with Prinz's view. In principle, I think embodied appraisals are compatible with CAT. For example, consider the following illustration from Prinz:

> Consider the chain of events leading to fear. Something dangerous occurs. That thing is perceived by the mind. The perception triggers a constellation of bodily changes. These changes are registered by a further state: a bodily perception. The bodily perception is directly caused by bodily changes, but it is indirectly caused by the danger that started the whole chain of events. It carries information about danger by responding to changes in the body. That further state is fear (2004a, pg.69).

CAT theorists can accept everything that Prinz says here with a caveat. Insofar as the subject engages in a conceptual act that results in their categorizing the experience as fear, "that further state is fear". Because Prinz is committed to basic emotion theory, he thinks that fear arises

[30] One might object here that categorization is different from an emotional experience, so CAT is mistaken. But this is to beg the question against CAT because that is exactly what CAT denies. It is only through categorizing your experience (or your perception of another) as a particular emotion (say, anger), that one experiences anger.

independent of categorization. But this is not the case for psychological constructionists, and CAT in particular (more on this below).

4.3 Part 2: Sensorimotor Reenactments are not Sufficient

As I reviewed in section 3.3, there is mounting evidence that emotions are not basic, discrete kinds. Instead, emotional experience and perception is highly variable both inter- and intra-personally across time. Moreover, the elements that make up a token emotional experience are many, including affect, facial muscle movements, vocal signals, body language, and context. The perceptual fuzziness of emotions has consequences for the human emotional conceptual system: "if the body and the situation help to constitute the mind, then the structure and content of the conceptual system for emotion should be grounded in the structure and content of emotional events as they naturally occur" (Barrett and Lindquist 2008, pg.17).

Because emotional experience and perception are fuzzy, emotional concepts will be too.[31] This is due to those concepts being embodied simulations of naturally occurring emotional experiences and perceptions. So just as experiencing anger on a particular occasion may involve, for example, shaking, furrowing the brow, shouting, and the urge to break stuff, activation of the concept "anger" might involve simulating these responses by subthreshold reactivation of the neural substrates that underlie those responses. I say "may" and "might involve" here because, as already noted, these responses can vary. If this is right, then sensorimotor reenactments will not be sufficient information to identify the target emotional category. Or, more cautiously, it might

[31] I actually think that this explanation runs in the other direction too: that because our concepts are fuzzy, emotional experience and perception is fuzzy too. To foreshadow the next subsection, labels serve to make our emotional concepts discrete as well as emotional perception. Lupyan (2012) calls this bi-directional linguistic modulation "the label feedback hypothesis".

be enough information, but there is too much information (e.g., too many perceptual properties or too many subtle variations between reenactments) to identify the target category.

Contrast this view with the basic emotion theorists' view of emotional concepts. The basic emotion theorist hypothesizes, for example, that there are one-to-one correspondences between facial muscle configurations and emotional categories. If this is true, then knowing those specific configurations (or simulating those configurations) will be sufficient to identify the target emotional category. If emotional experience and perception are discrete, then so too will be emotional concepts. But as I argued in section 3.3, this is not the case.

4.4 Part 3: Labels as Binders

In section 3.4 I discussed a finding from Lindquist and Gendron (2013) that showed that without labels, healthy participants range from 7.5% to 54% accurate at emotional recognition tasks. However, with labels, healthy participants jump to above 63% accurate. CAT hypothesizes that language serves as a glue to bind together the various bodily responses into discrete categories.[32] In particular, emotional labels (e.g., "anger", "fear", "happy") organize humans' emotional conceptual system by creating a kind of filing system (see Prinz 2005, pg. 683). When a child undergoes an emotional experience (e.g., cries because big brother took away her rattle) and the parent asks "Why are you sad?," the parent is labeling the child's very complex experience as an instance of "sadness". As Lindquist (2013) puts it:

> Indeed, emotion categories might be acquired in childhood by bootstrapping situations and core affective feelings to the words used by adult caregivers (e.g., when mom and

[32] A similar point is made by Prinz (2005): "verbal labels serve as placeholders for ideas that are too complex to hold in one's mind all at once" (pg. 692). Due to emotional experiences and perceptions being so complex and variable, labels simplify. But importantly, as Prinz notes, the label by itself is not enough. The label "must be pinned down to the senses in order to be applied in the world" (pg.692). Hence why the labels serve as file folders that contain embodied representations.

> dad tell Joey not to be "sad" because of a broken toy, Joey learns that negative feelings following a loss are associated with the category "sadness" in his culture) (pg.362).

As the child begins to make strong associations between situations, bodily responses, and emotional labels, the child begins a filing system with folders for each emotional category.[33] Inside these folders are packets of embodied simulations that are strongly tied to that emotion. For example, in the "sad" folder, there might be simulations for crying, hunching the shoulders, pain in the chest, frowning, etc.[34]

Moreover, as Clark (1998) and Lupyan et al. (2007) have argued, labels not only facilitate learning of categories, they also facilitate categorization even after that concept has been learned. The filing system metaphor will be helpful here. If one is trying to identify what emotion a person is experiencing, it helps when the conceptual landscape is organized into discrete folders with labels for each. When labels are provided on a task, those labels serve as direct routes to conceptual knowledge by activating embodied simulations. Consistent with the evidence canvassed in section 3, labels, according to CAT, bind together the highly variable instances of emotional categories, chunking them into discrete folders of embodied information. Sensorimotor reenactments are not sufficient for emotional recognition because language is needed to partition fuzzy emotional concepts into discrete concepts.

---

[33] Widen (2013) reviews data suggesting that children's emotional conceptual system begins with two broad categories: good and bad. As these children learn emotional labels, their emotional conceptual systems begin to sharpen into discrete categories. It appears that the maturation of the emotional conceptual system happens concurrently with the expansion of emotional vocabulary.

[34] Paul Pietroski (2005) makes a similar point about lexicalization (and its relation to concepts) in general saying that "lexicalization is a process in which diverse mental representations can be linked via the language system. Perhaps without lexicalization, representations that are different in kind cannot be combined to form a complex concept that is usable in human thought, but (luckily for us) the language system provides resources for creating certain ''common denominators'', which make it possible to create endlessly many complex mental representations with constituents that are typologically disparate" (pg. 271).

4.5 Explaining Moebius Patient Performance

CAT is an embodied, constructionist view of emotions. It says that emotions emerge in consciousness when a subject categorizes various bodily responses as a particular emotional category. Categorization, on this view, involves sensorimotor reenactments. But as we saw, sensorimotor reenactments will only get one so far. Because emotional categories are perceptually fuzzy and complex, language is needed to bind together the heterogeneous and highly variable bodily responses into perceptually discrete categories. Once the emotional categories are partitioned, labels further serve to organize the conceptual landscape into file folders containing simulations of each emotional category. Emotion labels then serve as direct routes to emotional conceptual knowledge by activating embodied simulations. Let us now turn to how CAT explains the mixed results of the Moebius syndrome studies.

I'll start by explaining Moebius' patient performance when labels are provided. The reader will recall that in these studies, Moebius patients performed comparably to healthy controls. According to CAT, when labels are provided on a task they serve as direct routes to emotional conceptual knowledge by activating embodied simulations. Of course, the patients cannot simulate facial expressions themselves; however, they have seen people make facial expressions in various contexts. So they might project themselves into a context (e.g., episodic simulation) wherein that facial expression was made.[35] Moreover, because labels bind together various bodily responses to discrete categories, labels don't just activate facial muscle simulations. They activate simulations for various bodily responses typical of the target emotional category. So perhaps label-activated simulations provide richer information than an

[35] For those skeptical that episodic simulation is embodied, Lindquist (2013, pg.362) explicitly mentions episodic memory as a form of sensorimotor reenactment. See also De Brigard (2014) and Schacter et al. (2015).

emotional face in isolation. Even though Moebius patients cannot make or simulate facial expressions, this does not preclude them from successful recognition. Embodied emotional concepts are not exhausted by facial muscle mimicry.

What does CAT say about Moebius patients' performance when labels are not provided? Recall that Moebius patients are severely deficient in these studies, so much so that the patients could not respond at all to the task. The patient in Giannini et al.'s (1984) study described the experience as like trying to understand a foreign language. Firstly, without labels provided, subjects do not know where in the conceptual landscape to look for the target emotion. This is true for both Moebius patients and healthy controls. But Moebius patients have an additional handicap which is that facial muscle mimicry is not part of the content of their emotional concepts. So the information that they are given (the presented picture of an emotional facial expression) is not a guide (in isolation) to the target category. Contrastingly, for healthy controls, even though they are not provided labels, they *can* mimic the facial expressions. So rather than having no guide as to what emotion is being depicted in the presented picture, they have a rough guide. This explanation of Moebius patients' performance vindicates the report from Giannini et al.'s patient because having no information to use in reading a facial expression is, plausibly, very much like trying to understand a foreign language.

5. Limitations and Shortcomings

Before concluding the paper, I want to address some limitations of the analysis provided here. Firstly, I showed that there is a pattern in the empirical literature on Moebius patient performance on face-based emotion recognition tasks that is attributable to differences in experimental design. The pattern was that when patients are provided labels in the task, they

perform comparably to controls, but when they are not provided labels, the patients are impaired. There is an exception to this trend that needs to be addressed.

Bate et al. (2013) did a face-based emotion recognition study with Moebius patients (n=6). In this study, 5 out of the 6 patients were impaired *even when emotion labels were provided*. This finding is in tension with not only the trend I highlighted, but also with the larger theoretical point I want to make: that labels allow the patients to perform comparably to controls by activating embodied conceptual knowledge (sans facial muscle mimicry).[36]

However, Bate et al. (2013) deployed an experimental design that involves time restriction on exposure to the target emotional face. Specifically, "stimuli are presented in a random order for five seconds per face, followed by a blank screen" (pg.e62656). After the blank screen, subjects must choose between the six basic emotions. The selection portion of the task is not timed. Here, subjects hold the presented face in working memory, and then select the correct emotion label. This difference in experimental design adds computational costs to the task. For one, it is well known that storage in working memory requires attentional resources (Cowan 2005). And secondly, representations in working memory are disposed to decay due to issues in encoding (Baddeley and Hitch 1974). If, in the absence of labels, Moebius patients must rely on facial expressions to identify the target emotion, and they cannot simulate facial expressions, then it's not clear that patients would be able to hold an embodied representation of an emotional face in working memory-- especially not the details of the facial expression. This is conjecture, but I suspect that encoding of the facial expression among Moebius patients is degraded due to their facial paralysis.

---

[36] Thanks to an anonymous reviewer for this!

Some support for this conjecture actually comes from the Warrington memory task in Calder et al 's (2000). In this task, subjects are shown fifty faces individually for approximately three seconds each. During those three seconds, subjects are to report whether the face looks pleasant or unpleasant. Then, they are given fifty pairs of faces (in each pair, one is from the training set and one is *not* from the training set). Subjects are asked to identify the face from the training set. The patient mean score for this task was 36.67/50.00 whereas the control mean score was 43.60/50.00. The significant impairment on this task suggests that Moebius patients might have encoding impairments for emotional faces.

More straightforward evidence for the claim that Moebius patients have difficulty encoding representations of emotional faces in visual working memory comes from Gambarota et al. (2020). These authors issued a delayed estimation task with no verbal component. Here, subjects were presented with a to-be-memorized emotional face of a certain intensity for approximately 1 second. They were then shown an array of emotional faces and instructed to select the face they were initially presented with. Gambarota et al. reported a significant difference between Moebius subjects and controls, with Moebius subjects deficient. The authors conclude that Moebius "participants built lower quality representations of the intensity of emotional expressions when compared to healthy participants. These findings support the role of sensorimotor simulation in improving the quality of emotional representations of facial expressions during early stages of processing" (pg. 2).

One final concern I'd like to address came from an anonymous reviewer. The concern is that it's possible that labels are *not* cues to embodied simulations at all. Rather, it could be that labels allow the subjects to guess among the choices provided. For example, if provided no labels the task may be more difficult than if 4 labels were provided simply because in the latter

case one has a 25% chance of getting it right by guessing. If one is given 2 labels, it's even easier. In other words, the objection agrees that emotions are perceptually fuzzy, but disagrees that emotions are embodied. So all that labels do is reduce the search space for the relevant emotional category.

There seems to be two subtly different objections here that I'm going to separate out. The first is that Moebius subjects are guessing when labels are provided. The second is that there is a label superiority effect, but that the explanation for the superior performance is not an embodied one. The former objection is consistent with embodiment because guessing when labels are provided does not presuppose nor entail that embodiment (in the sense I have defined it here) is false. For example, labels might activate embodied simulations, but *only* facial muscle simulations. In this case, Moebius patients would not be helped by labels being provided, so they guess among the options provided by the labels. The latter objection outright denies that labels activate *any* embodied simulations, but grants my arguments for the perceptual fuzziness of emotional categories. The implication for Moebius' patient performance is that provided labels reduce the search space of emotional categories, facilitating categorization, but not via embodied simulations. I'll respond to the guessing objection first, then handle the reducing the search space objection.

If it's possible to guess when labels are given, it's also possible to guess when labels are not given. Afterall, Moebius patients have an emotional vocabulary just like controls do. However, as reviewed, when labels are *not* provided Moebius patients fail to respond to the task. This suggests that guessing is not at play here.

Still, the objector could grant this point and suggest that a forced-choice paradigm makes it *more likely* that subjects would guess. This is certainly possible; however, Giannini et al.

(1984) used labels for amounts of winnings that the subject was instructed to select on the basis of video clips of people's emotional reactions at a slot machine. As reported, the subject failed to respond to the task. Granted, this is a sample size of 1, so it's hard to generalize. But the point is that since the subject did not guess here, why think they are guessing in the other experiments?

However, I grant that it is certainly possible that subjects are guessing in these tasks. I can't rule that out, but I think it would be surprising. Here's why: In the studies in which controls and Moebius subjects were provided labels to match to emotional faces, both groups performed comparably. Suppose that both groups were guessing in these tasks. It would be incredibly coincidental that they ended up performing comparably to each other. Further, suppose only the Moebius subjects were guessing–the controls were not. Still, it would be incredibly coincidental that patient performance was comparable to the controls. On the other hand, if you buy my explanation, it is not coincidental at all. Emotion labels serve as direct routes to emotional concept knowledge through activating embodied simulations in both groups.

Now for the second objection. Is it possible that the label superiority effect, in the face-based emotion recognition tasks, is best explained by a non-embodied reduction of the search space of emotional categories? At the end of section 3, I considered this explanation for the label superiority effect. There, I said that it's not clear how the label superiority effect helps the embodied theorist because labels, *prima facie*, activate non-embodied, amodal representations. Thus, it is not clear how this shows that Adams' premise 1 is false. Indeed, you might think that the label superiority effect supports Adams' premise 1. In responding to this, I showed how the CAT accommodates the label superiority effect. However, you might be unsatisfied with this response because this only shows that we have two competing explanations of Moebius patient

performance– one that is embodied and one that is not. Therefore, we need positive reasons for thinking that language, and in particular labels, activate embodied simulations.

The best evidence for my case would likely be imaging data on Moebius patients and controls during an emotional recognition task showing sensorimotor activation when labels are provided. At the time of writing this, I know of two imaging studies with Moebius patients (Japee et al. 2022; Sessa et al. 2022), but neither used emotion labels. Additionally, both of these studies deployed a match-to-sample paradigm. While interesting, this paradigm does not probe the target phenomenon– emotion labels'(in)ability to induce embodied simulations. The lack of evidence however is not evidence of a lack. This is an area where further work should be done.

However, there is positive evidence for embodied simulations in language comprehension more generally. As briefly mentioned in the introduction, the action-sentence compatibility effect is a well established finding (Bergen 2012). There is also evidence of embodied simulation in metaphor comprehension (Carston 2010; Gibbs, Jr. 2006; Green 2017; Lakoff and Johnson 2008), dialogue processing (Pickering and Garrod 2004; Zwaan and Radvansky 1998), and grammar processing (Bergen 2012).

I think most relevant to present purposes is recent work on polysemy resolution. Polysemy can be understood as an instance of lexical ambiguity in which "words have a single meaning that can be modulated to fix distinct denotations depending on context" (Quilty-Dunn 2021; pg.164). For example, "Jack chugged the can, and crushed it on his head." has two occurrences of "can". In the first instance, "can" denotes the liquid that Jack chugged, but in the second instance "can" denotes the smashed aluminum object. While this sentence seems felicitous, some sentences involving polysemes are infelicitous: "This chicken is scrumptious

and chirpy!". Here, the first occurrence of "chicken" denotes the food item while the second denotes the farm animal.

Michelle Liu (forthcoming) argues that polysemy resolution depends upon embodied simulations. In particular, she argues that infelicitous polysemy-involving sentences entail conflicting simulations whereas felicitous ones do not. The chicken sentence entails conflicting simulations because it involves the comprehender simulating chicken-as-meat and chicken-as-animal, but our everyday experiences with chicken meat do not involve it being chirpy. Contrastingly, in the Jack sentence, can-as-liquid and can-as-vessel overlap in our everyday experiences. This sentence does not involve conflicting simulations because it does not demand the comprehender to simulate a fundamentally different entity (as in the chicken sentence).

I have argued that emotional categories are perceptually fuzzy and complex because emotional experience and perception are highly variable inter- and intra-personally through time. Emotion labels are needed, according to CAT, to bind together those heterogenous representations into unified wholes. This means that "angry" can be used to denote, for example, a red-faced person shouting and throwing furniture, but also a person quietly standing in a corner with their arms crossed. Emotion labels can be used to denote different objects depending on context. If the best explanation of polysemy resolution involves embodied simulations, then we should expect that comprehending emotional labels also involve embodied simulations. Again, though, more work should be done on this front.

5. Conclusion

Fred Adam's argument against embodied concepts has been shown to be unsound. Premise 2, which states that Moebius patients succeed on emotional recognition tasks has been shown to be misleading. These patients sometimes succeed on such tasks, and when they do they

are provided emotional labels to match to the facial expression. When these patients are not provided labels, they fail to respond to the task. In explaining these performance differences by reference to psychological constructionism, I have shown that Adam's premise 1 is false. It is not the case that if Moebius patients succeed on face-based emotional recognition tasks, then embodiment is false. Indeed, CAT, as I showed, is an embodied view of emotions. Because CAT can explain Moebius patients' successful performance, embodiment is not threatened. In fact, embodied views, and in particular CAT, gains motivation.

Besides defending the embodiment of emotion concepts, we have learned some things about the nature of emotion and the language-cognition-perception interface. I discussed overwhelming evidence that shows that, contra basic theories, emotions are constructed out of various more basic psychological components. For this reason, emotions are perceptually fuzzy and complex. I also discussed how perceptually fuzzy categories are (at least) easier to learn with the help of labels. Moreover, once those concepts are learned, labels facilitate categorization by simplifying the environment and organizing our conceptual landscape. Despite common sense dictating that emotions are innate, pre-linguistic, and subserved by distinct mechanisms, emotions are actively (albeit implicitly) constructed, via conceptual acts, for the context that the subject finds themself in.

REFERENCES

Adams, F. (2010). Embodied cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 619-628.

Allport, F. H. (1924). *Social psychology*. Boston, Houghton.

Baddeley, A., & Hitch, G. (1974). Working Memory: In Bower GA. The psychology of learning and cognition.

Barrett, L. F. (2006). Are emotions natural kinds?. *Perspectives on psychological science*, *1*(1), 28-58.

Barrett, L. F., & Lindquist, K. A. (2008). The embodiment of emotion. *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*, 237-262.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(4), 577-660.

Bate, S., Cook, S. J., Mole, J., & Cole, J. (2013). First report of generalized face processing difficulties in möbius sequence. *PloS one*, *8*(4), e62656.

Bell, C., Nevitt, S., McKay, V. H., & Fattah, A. Y. (2019). Will the real Moebius syndrome please stand up? A systematic review of the literature and statistical cluster analysis of clinical features. *American Journal of Medical Genetics Part A*, *179*(2), 257-265.

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. Basic Books.

Bergen, B., & Feldman, J. (2008). Embodied concept learning. In *Handbook of Cognitive Science* (pp. 313-331). Elsevier.

Bertoux, M., Duclos, H., Caillaud, M., Segobin, S., Merck, C., de La Sayette, V., ... & Laisney, M. (2020). When affect overlaps with concept: emotion recognition in semantic variant of primary progressive aphasia. *Brain*, *143*(12), 3850-3864.

Blouw, P., Solodkin, E., Thagard, P., & Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive science*, *40*(5), 1128-1162.

Bogart, K., & Matsumoto, D. (2010). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social neuroscience*, *5*(2), 241-251.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology.

Buckner, C. (2018). Empiricism without magic: Transformational abstraction in deep convolutional neural networks. *Synthese*, *195*(12), 5339-5372.

Calder, A. J., Keane, J., Cole, J., Campbell, R., & Young, A. W. (2000). Facial expression recognition by people with Möbius syndrome. *Cognitive neuropsychology*, *17*(1-3), 73-87.

Caramazza, A., & Mahon, B. Z. (2006). The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology*, *23*(1), 13-38.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and brain sciences*, *25*(6), 657-674.

Carston, R. (2010, October). XIII—Metaphor: Ad hoc concepts, literal meaning and mental images. In *Proceedings of the Aristotelian society* (Vol. 110, No. 3_pt_3, pp. 295-321). Oxford, UK: Oxford University Press.

Casasanto, D., & Lupyan, G. (2011). Ad hoc cognition. In *The 33rd Annual Conference of the Cognitive Science Society [CogSci 2011)* (p. 826). Cognitive Science Society.

Chemero, A. (2011). *Radical Embodied Cognitive Science*. MIT Press.

Clark, A. (1998). Magic words: how language augments. *Language and Thought: Interdisciplinary Themes*, 162.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, *58*(1), 7-19.

Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind*. MIT press.

Cowan, N. (2005). Working memory capacity limits in a theoretical context. In *Human learning and memory: Advances in theory and application. The 4th Tsukuba international conference on memory* (pp. 155-175). Mahwah, NJ: Erlbaum.

Darwin, C. (1872). *The Expression of Emotions in Man and Animals.* Murray.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, *191*(2), 155-185.

De Stefani, E., Ardizzi, M., Nicolini, Y., Belluardo, M., Barbot, A., Bertolini, C., ... & Ferrari, P. F. (2019). Children with facial paralysis due to Moebius syndrome exhibit reduced autonomic modulation during emotion processing. *Journal of Neurodevelopmental Disorders, 11*(1), 1-16.

De Stefani, E., Nicolini, Y., Belluardo, M., & Ferrari, P. F. (2019). Congenital facial palsy and emotion processing: The case of Moebius syndrome. *Genes, Brain and Behavior*, *18*(1), e12548.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain* (No. 202). WW Norton & Company.

Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. *Minds, machines and evolution*, 129-151.

Dennett, D. C. (2015). *Elbow Room, new edition: The Varieties of Free Will Worth Wanting*. mit Press.

Deonna, J., & Teroni, F. (2012). *The emotions: A philosophical introduction*. Routledge.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, *11*(8), 333-341.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412-431.

Dove, G. (2011). On the need for embodied and dis-embodied cognition. *Frontiers in Psychology*, *1*, 242.

Edelman, S., & Intrator, N. (1997). Learning as formation of low-dimensional representation spaces. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, Erlbaum, Mahwah, NJ* (pp. 199-204).

Edelman, S., & Intrator, N. (1997). Learning as extraction of low-dimensional representations. *Psychology of learning and motivation*, *36*, 353-380.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion review*, *3*(4), 364-370.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, *1*(1), 56-75.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive science*, *33*(4), 547-582.

Farina, M. (2021). Embodied cognition: dimensions, domains and applications. *Adaptive Behavior*, *29*(1), 73-88.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.

Gallagher, S. (2011). Interpretations of embodied cognition. In W. Tschacher and C. Bergomi (eds.), *The Implications of Embodiment: Cognition and Communication*. Exeter: Imprint Academic.

Gambarota, F., Luria, R., Pastore, M., Ferrari, P. F., & Sessa, P. (2020). Deficits of visual working memory representations of emotional facial expressions in patients with congenital facial palsy. *psyarXiv* 2020.

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, *12*(2), 314.

Giannini, A. J., Tamulonis, D., Giannini, M. C., Loiselle, R. H., & Spirtos, G. (1984). Defective response to social cues in Möbius' syndrome. *Journal of Nervous and Mental Disease*.

Gibbs Jr, R. W. (2006). Metaphor interpretation as embodied simulation. *Mind & Language*, *21*(3), 434-458.

Green, E. J. (2020). The perception-cognition border: A case for architectural division. *Philosophical Review*, *129*(3), 323-393.

Green, M. (2017). Imagery, expression, and metaphor. *Philosophical Studies*, *174*(1), 33-46.

Gregor, K., Besse, F., Jimenez Rezende, D., Danihelka, I., & Wierstra, D. (2016). Towards conceptual compression. *Advances In Neural Information Processing Systems*, *29*.

Griffiths, P. E. (2008). *What emotions really are: The problem of psychological categories*. University of Chicago Press.

Huang, L., & Awh, E. (2018). Chunking in working memory via content-free labels. *Scientific reports*,

8(1), 23.

James, W. (1884). What is an emotion? *Mind, 9,* 188-205.

Japee, S., Jordan, J., Licht, J., Lokey, S., Moebius Syndrome Research Consortium, Chen, G., ... & Ungerleider, L. G. (2022). Inability to make facial expressions dampens emotion perception. *bioRxiv*, 2022-10.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*(2), 113-139.

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. *Psychological stress: Issues in research*, 14-37.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120392.

Lange, C. G. (1885). *Om sindsbevaegelser: et psyko-fysiologisk studie.* Kjbenhavn: Jacob Lunds. Reprinted in *The emotions*, C. G. Lange and W. James (eds.), I. A. Haupt (trans.) Baltimore, Williams & Wilkins Company 1922.

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, *5*(4), 356-368.

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological science*, *19*(9), 898-903.

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66-71.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion*, *14*(2), 375.

Liu, M. (forthcoming). Mental simulation and language comprehension: The case of copredication. *Mind & Language.* https://philarchive.org/archive/LIUMSA-2

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in psychology*, *3*, 54.

Lupyan, G., & Bergen, B. (2016). How language programs the mind. *Topics in cognitive science*, *8*(2), 408-424.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279-284.

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1319-1337.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by

verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in cognitive sciences*, *24*(11), 930-944.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological science*, *18*(12), 1077-1083.

Mahon, B. Z. (2015). What is embodied about cognition?. *Language, cognition and neuroscience*, *30*(4), 420-429.

Matsumoto, D., & Ekman, P. (2004). The relationship among expressions, labels, and descriptions of contempt. *Journal of personality and social psychology*, *87*(4), 529.

Matsumoto, D., & Ekman, P. (2006). Multi-ethnic facial expression set. *Unpublished images*.

McClamrock, R. (1995). *Existential cognition: Computational minds in the world*. University of Chicago Press.

Michael, J., Bogart, K., Tylén, K., Krueger, J., Bech, M., Østergaard, J. R., & Fusaroli, R. (2015). Training in compensatory strategies enhances rapport in interactions involving people with Möbius syndrome. *Frontiers in Neurology*, *6*, 213.

Michel, C. (2021). Overcoming the modal/amodal dichotomy of concepts. *Phenomenology and the Cognitive Sciences*, *20*(4), 655-677.

Möbius, P. J. (1888). Ueber angeborene doppelseitige Abducens-Facialis-Lahmung. *Munch med wochenschr*, *35*, 91-94.

Newen, A., De Bruin, L., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

Nicolini, Y., Manini, B., De Stefani, E., Coudé, G., Cardone, D., Barbot, A., ... & Ferrari, P. F. (2019). Autonomic responses to emotional stimuli in children affected by facial palsy: The case of Moebius syndrome. *Neural plasticity*.

Núñez, R. (2008). Conceptual metaphor, human cognition, and the nature of mathematics. *The Cambridge handbook of metaphor and thought*, 339-362.

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: do animals have affective lives?. *Neuroscience & Biobehavioral Reviews*, *35*(9), 1791-1804.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66-72.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Pietroski, P. (2005). Meaning before truth. *Contextualism in philosophy: Knowledge, meaning, and truth*, *255*, 302.

Pons, F., Harris, P. L., & Doudin, P. A. (2002). Teaching emotion understanding. *European Journal of Psychology of Education*, *17*(3), 293-304.

Prinz, J. J. (2004a). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Prinz, J. J. (2004b). Emotions embodied. *Thinking about feeling: Contemporary philosophers on emotions*, 44-59.

Prinz, J. J. (2004c). *Furnishing the mind: Concepts and their perceptual basis*. MIT press.

Prinz, J. J. (2005). The return of concept empiricism. In *Handbook of categorization in cognitive science* (pp. 679-695). Elsevier Science Ltd.

Quilty-Dunn, J. (2021). Polysemy and thought: Toward a generative theory of concepts. *Mind & Language*, *36*(1), 158-185.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive science*, *34*(6), 909-957.

Rasmussen, L. K., Rian, O., Korshoej, A. R., & Christensen, S. (2015). Fatal complications during anaesthesia in Moebius syndrome: a case report and brief discussion of relevant precautions and preoperative assessments. *International Journal of Anesthesiology & Research*, *3*(6), 116-118.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145.

Russell, J. A., & Widen, S. C. (2002). A label superiority effect in children's categorization of facial expressions. *Social Development*, *11*(1), 30-52.

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review*, *3*(4), 444-454.

Schacter, D. L., Benoit, R. G., De Brigard, F., & Szpunar, K. K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of learning and memory*, *117*, 14-21.

Sessa, P., Schiano Lomoriello, A., Duma, G. M., Mento, G., De Stefani, E., & Ferrari, P. F. (2022). Degenerate pathway for processing smile and other emotional expressions in congenital facial palsy: an hdEEG investigation. *Philosophical Transactions of the Royal Society B*, *377*(1863), 20210190.

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual review of psychology*, *54*(1), 145-173.

Shapiro, L. (2019). *Embodied cognition*. Routledge.

Shariff, A. F., & Tracy, J. L. (2011). Emotion expressions: On signals, symbols, and spandrels—A response to Barrett (2011). *Current Directions in Psychological Science*, *20*(6), 407-408.

Shweder, R. A. (1994). You're not sick, you're just in love": Emotion as an interpretive system. *The nature of emotion: Fundamental questions*, 32-44.

Shweder, R. A., Haidt, J., Horton, R., & Joseph, C. (1993). The cultural psychology of the emotions. *Handbook of emotions*, 417-431.

Slaby, J. (2014). Emotions and the extended mind. *Collective emotions*, 32-46.

Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In *Advances in psychology* (Vol. 93, pp. 233-272). North-Holland.

Stewart, T. C., Bekolay, T., & Eliasmith, C. (2011). Neural representations of compositional structures: Representing and manipulating vector spaces with spiking neurons. *Connection Science*, *23*(2), 145-153.

Valeriani, L., Doimo, D., Cuturello, F., Laio, A., Ansuini, A., & Cazzaniga, A. (2023). The geometry of hidden representations of large transformer models. *arXiv preprint arXiv:2302.00294*.

Vannuscorps, G., Andres, M., & Caramazza, A. (2020). Efficient recognition of facial expressions does not require motor simulation. *Elife*, *9*, e54687.

von Graefe A. In: von Graefe A, Saemisch T, eds. Handbuch Der Gesamten Augen-heilkunde. Vol 6. Leipzig, Germany: Engelmann; 1880:60-67.

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*(1), 72-77.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.

Chapter 2: What's in a Color?: Labels, Gesture, and Synesthesia

1. Introduction

Does learning a natural language change the way we see and think about the world? There's been a wealth of recent research suggesting exactly this. A paradigm case is the so-called "label superiority effect" (Russell and Widen 2002). For example, subjects trained on novel stimuli with labels are faster and more accurate at identifying novel exemplars than subjects trained without labels, even when the labels are nonsense (Lupyan et al. 2007). One commonly accepted explanation for this effect is that language makes perception more "categorical" which facilitates lexical access (for reviews see Lupyan 2012; Lupyan et al. 2020). Perception is more categorical when between-category differences are more salient than within-category differences, as when the boundary between red and yellow is more salient than reddish-orange and reddish-yellow (Holmes and Reiger 2017; Reiger and Xu 2017).

Is language special in its ability to make perception more categorical? Recent work on gesture seems to suggest otherwise. Gestures have been shown to facilitate performance in a variety of domains including mathematics (Khatin-Zadeh 2022), communication (Hostetter 2011), generalization (Goldin-Meadow 2015), and spatial cognition and navigation (Delgado et al. 2011; So et al. 2014). One of the most commonly accepted explanations for gestures' cognitive advantages is that gestures "schematize" representations (Goldin-Meadow 2015). Schematization is a kind of abstraction wherein certain irrelevant details of a representation are deleted or ignored (Khatin-Zadeh et al. 2023). This allows the subject to focus attention to the most relevant features of a task while disregarding irrelevant features.

The similarity between the explanations for language's ability to make perception more categorical and gestures' schematization is striking. They both seem to involve noise reduction in

an effort to simplify representations. The idea is that without gestures or language, representations of some target domain can be too perceptually complex for efficient and accurate action (e.g., problem solving, categorization, learning). Language and gesture are both forms of representation that simplify the task environment by directing attention to the most task relevant features and reducing noise.

In this paper, I want to push this explanation further. Recent work on synesthesia has shown that synesthetes have cognitive advantages owing to their unique experiences of the world. For example, grapheme-color synesthetes see letters and words as colored (regardless of what color the letters and words are printed in). These synesthetes have superior memory performance for recall of word lists (Radvansky et al. 2011). I am going to focus on a recent study by Speed and Majid (2018) showing that olfactory-color synesthetes have superior categorization performance on smell identification tasks. Olfactory-color synesthetes experience smells as colored– they "see" smells. I will argue that their superior performance on categorization tasks is due to information reduction of olfactory representations by automatically elicited color representations. In other words, just like gestures and language, these color representations make olfactory representations more categorical, facilitating lexical access.

The plan for the paper is as follows: in section 2, I will be reviewing work on the label superiority effect in categorization and identification tasks. In section 3, I will review work on gesture and schematization. I will end this section by arguing that both language and gesture facilitate cognitive performance by making representations more categorical through noise reduction. I will then turn to olfactory-color synesthesia, in section 4, arguing that the synesthetic color associations also make representations more categorical through noise reduction. Here, it will be established that language is not special in its ability to make representations more

categorical. Indeed, I will argue that there is likely a common mechanism for linguistic, gestural, and synesthetic categorical effects on representations. Finally, in section 5, I will consider objections.

2.The Label Superiority Effect

Categorization and identification tasks are induction tasks. When a subject is presented with a token instance of a category, she has to infer what category this token belongs to. Categorization can be as simple as sorting various stimuli into groups (e.g., all the red items here, and the blue items there). It needn't involve language. However, when a subject is tasked with identifying the category that the stimulus belongs to, language enters the fold. The subject must connect their perceptual representation of the stimulus to a linguistic representation of the category. In this section, I will briefly review evidence that language intervenes on categorization. In particular, having access to verbal labels of categories makes discrimination and categorization easier.[37]

To pump the reader's intuitions, have a look at the following images of butterflies:

PICTURE 3. https://kids.nationalgeographic.com/animals/invertebrates/facts/monarch-butterfly

[37] There is additional evidence that language can modulate discrimination and even detection (see Lupyan et al. 2020).

PICTURE 4. https://en.wikipedia.org/wiki/Viceroy_%28butterfly%29

Now, take a look at this third butterfly and consider which it is most similar to:

PICTURE 5. https://www.imagineourflorida.org/viceroy-butterfly/

Hopefully, you think that the third butterfly is most similar to the second. The first butterfly is a Monarch, but the second and the third are Viceroy butterflies. Before reading further, look back at the butterflies and see if you can spot the differences.

There are a few subtle differences between Monarchs and Viceroys, but probably the most prominent difference is the black markings on the hindwings. Viceroys have a horizontal-ish black stripe that runs through the vertical stripes, whereas Monarchs have a horseshoe shaped black stripe that

connects to their vertical stripes. Chances are, you could more easily tell these butterflies apart after I assigned them their category labels. This phenomenon is known as “the label superiority effect” (Russell and Widen 2002) and it is a well-established finding in cognitive psychology.

Perhaps the most well-known instance of the label superiority effect comes from color categorization tasks with Russian and English speakers (Winawer et al. 2007). Russian and English divide the color spectrum differently. Russian divides blue into two different categories (with no superordinate label corresponding to English “Blue”)– “goluboy” for lighter blues and “siniy” for darker blues. Winawer et al. tested English and Russian speakers in a speeded discrimination task with blue stimuli. The results indicated that Russian speakers were faster at discriminating darker blues from lighter blues than English speakers, but they were slower than English speakers when the stimuli were both dark or both light blue. These results suggest that labels enhance discrimination performance for between-category stimuli, but inhibit performance for within-category stimuli. Consistent with the definition of categorical perception–that between-category differences are more salient than within-category differences– it would seem that the Russian speakers benefitted from the label superiority effect precisely because their perception was more categorical.

Interestingly, Winawer et al. (2007) supplied verbal interference in a separate experiment with the same methods and stimuli, and the effect on Russian speakers vanished. This is interesting because it suggests that the label superiority effect is an on-line process. That is, the induction of more categorical perception occurs during the task. It is not the case that the Russian speakers’ representations of dark and light blues are stored as more categorical, rather they become more categorical via online linguistic modulation.

In a landmark study by Lupyan et al. (2007), subjects were faster and more accurate at categorizing novel stimuli when provided with a nonsense label than subjects not provided a label. Lupyan et al. trained subjects to associate two kinds of "aliens" with either of two behavioral responses: *approach* or *flee*. The two kinds of aliens were subtly different in shape (see figure 6). There were two training conditions: a label condition, in which the presented alien was labeled (either "leebish" or "grecious") and a no-label condition. Subjects were given feedback after each answer. After training, subjects were tested on how well they learned the associations. The results showed that subjects in the label condition, despite having the same amount of experience with the stimuli as subjects in the no-label condition, were faster and more accurate in both the training phase and at test. Lupyan et al. conclude that "learning nonsense verbal labels facilitated categorization more than did learning nonverbal associations" (2007, pg.1081).

PICTURE 6. The aliens on the left have flatter bases and a ridge on their head while the aliens on the right have rounder bases and no ridge (Lupyan et al. 2007, pg. 1078).

One final example. Perez-Gay Juarez et al. (2017) measured the induction of categorical perception in humans on images of black and white textures. There were two categories: Kalamite and Lakamite. The training phase involved selecting the label that applied to each image and then receiving corrective feedback. To learn these categories is to learn what

feature(s) were unique to each category. Consistent with the Winawer et al. study, the researchers found that upon learning the categories, human subjects perceived the textures more categorically– they perceived the between-category differences as more salient than the within-category differences. Perez-Gay Juarez et al. then modeled this effect in an auto-encoder network. The network matched human performance and this allowed the researchers to analyze categorical perception in an auto-encoder. The researchers concluded that the network was selecting the category relevant features and abstracting away from the irrelevant features– the network was reducing noise in its representation of the categories (Perez-Gay Juarez et al. 2017, pg. 23). If this is a good model of the human case, then perceiving the between-category differences as more salient than the within-category differences is (at least partly) explained by noise reduction. In particular, labels facilitate a reduction of category irrelevant information.

This explanation for categorical perception is echoed by Lupyan et al. (2007). In discussing the label superiority effect in the alien categorization task (reviewed above), they say the following:

> [R]ather than being fixed features, category names modulate item representations on-line through top-down feedback. According to this account, as a label is paired with individual exemplars, it becomes associated with features most reliably associated with the category. When activated, it then dynamically creates a more robust category attractor (pg.1082).

Lupyan (2012) further develops this view of the label superiority effect. In particular, he explains what it means for a representation to be a "more robust category attractor" with the *label feedback hypothesis*. This hypothesis states that:

> [L]anguage produces *transient* modulation of ongoing perceptual (and higher-level) processing. In the case of color, this means that after learning that certain colors are called "green," the perceptual representations activated by a green-colored object become

> warped by top-down feedback as the verbal label "green" is co-activated…Knowing that some colors are called green means that our everyday experiences of seeing become affected by the verbal term, which in turn makes the visual representation more categorical (pg. 4; emphasis original).

In being more categorical, Lupyan means that labels "help pull apart the representations [(of, say green and blue)] [which results] in decreased representational overlap between the two classes of stimuli" (pg.4). Lupyan thinks that this results in a reduction of category irrelevant information– when representations are "warped" by category labels to be more categorical, the category irrelevant information is abstracted away (see Goldstone and Hendrickson 2009; Goldstone et al. 2001; Lupyan and Lewis 2019; Lupyan and Thomspon-Schill 2012; Perry and Lupyan 2014). In the next section, I will be reviewing work on gestures showing similar cognitive advantages, and I will argue that gesture is also in the business of reducing category irrelevant information and noise.

3. Gestural Schematization as Reduction of Noise

It might be tempting to think that language is special in its ability to make representations more categorical. One might think this because category labels are discrete representations. The idea is that perceptual representations of categories are fuzzy and perpetually complex, but if you have access to a system of discrete symbols, you can reduce the complexity of perceptual representations. Some neo-empiricists have taken up this line (Clark 2006, 2012; Dove 2009, 2020, 2022; Gentry forthcoming). For example, Andy Clark (2006) has argued that we developed public language to augment basic biological thought. Basic biological thought is messy, driven by multi-modal sensorimotor learning, but access to a "public system of essentially context-free, arbitrary symbols" provides basic thought with the resources to "push,

pull, tweak, cajole and eventually cooperate with various non-arbitrary, modality-rich, context-sensitive forms of biologically basic encoding" (pg.5).

We should ask, however, whether it's the *format* of the representation that's doing the heavy lifting in augmenting perceptual representations, or whether it's sufficient to have access to a second representation, regardless of its format. In this section, I will review evidence that analog gestures can augment perceptual representations in the same way that language does–by reducing category irrelevant information.

When I lecture to my students, I am constantly waving my hands and arms around. I can't help but make these movements, but these movements are not meaningless reflexes, they are gestures. Gestures are a special type of sign that are ubiquitous in everyday communication. People use gestures to emphasize spoken words and highlight referents in the environment. Singer and Goldin-Meadow (2005) showed that gestures from a teacher can facilitate learning in students. Indeed, they showed that it can be a crucial difference maker as to whether students learn *at all*. The researchers tasked the teacher with teaching fourth graders mathematical equivalence (e.g., 7+6+4=4+__?). There were two conditions: a speech only condition (which included two different strategies for solving the problem) and a speech+gesture condition. In the speech+gesture condition, there were two sub-conditions: speech+gesture match and speech+gesture mismatch. In the mismatch condition, the teacher gives instructions in speech according to one strategy, but gestures in accordance with a different strategy. As Goldin-Meadow (2015) describes it:

> [T]he experimenter is teaching the problem 7 + 8 + 5 = __+5 and gives the same equalizer strategy in speech, adjusting for the different numbers; she said, "The way to solve this

> problem is to add 7 plus 8 plus 5 which equals 20; we want to make the other side of the equal sign the same amount; 15 plus 5 also equals 20 so 15 would be the answer." However, she produced an add–subtract strategy in gesture – she pointed at the 7, the 8, and the 5 on the left side of the equation and then produced a "take-away" gesture under the 5 on the right side of the equation, thus indicating in gesture that the numbers on the left should be added together and the number on the right subtracted from that sum (pg. 174).

In the speech+gesture match condition, the strategy was the same between speech and gesture. Singer and Goldin-Meadow (2005) found that children in the speech+gesture conditions solved significantly more problems correctly than the children in the speech only conditions. Indeed, children in the equalizer, speech only condition solved *no* problems correctly. Strikingly, even when the spoken instructions mismatched the gesture, children solved more problems correctly than the speech only conditions. Singer and Goldin-Meadow (2005) conclude that having access to two kinds of representation promotes learning better than one kind of representation.

The above case illustrates that when a teacher (or some interlocutor) produces gestures, it can promote learning in the listener. There is evidence that gestures can facilitate cognizing in the producer. Indeed, Khatin-Zadeh et al. (2022) note that gestures help the producer by, "priming and facilitating lexical retrieval, by maintaining visuospatial imagery, by organizing and encoding spatio-motoric information into suitable units, and by reducing the load of cognitive processing during speaking" (pg. 951).

However, all of this data is consistent with the hypothesis that a discrete representational format is necessary to boost cognitive performance. In other words, this data fails to show that language is not special in its ability to enhance performance. Indeed, gestures' cognitive

advantages might be explained by the combination of a discrete representational format (language) together with analog gesture. Goldin-Meadow (2015) puts the problem this way:

> Why does gesture promote learning? One possibility is that making use of two modalities at the same time strengthens the learner's representations of the problem, which, in turn, helps the learner take advantage of relevant input. Another possibility, however, is that it is not juxtaposing two modalities *per se*, but rather juxtaposing two different types of representational formats – a discrete categorical format found in speech and an analog mimetic format found in gesture…that promotes learning (pg. 177; emphasis original).

I interpret Goldin-Meadow as putting forward two competing hypotheses about how gesture might promote learning:

1. *Format Specific Thesis*: A necessary condition on gesture promoting learning is that it be combined with a discrete representation (e.g., language).
2. *Format Agnostic Thesis*: It is not necessary that gesture be combined with a discrete representation to promote learning.

To decide between these two hypotheses, Goldin-Meadow et al. (2012) look to deaf ASL-signers. Deaf signers, in addition to signing with their hands, also gesture with their hands. In mathematical equivalence tasks, the deaf signers used gestures approximately on a par with hearing children, but moreover the deaf children patterned the same performance as hearing children. From this data, one might reason that a discrete format of representation is not necessary for gesture to yield cognitive advantages. So the *Format Specific Thesis* is false, and the *Format Agnostic Thesis* is true.

This reasoning is too hasty because it assumes that ASL is *not* a discrete representational system. If it's true that ASL is not discrete, then this reasoning is valid. However, this assumption is controversial (see Goldin-Meadow and Brentari 2017). Hence, we still have not shown that language is not special in its ability to enhance performance on cognitive tasks, or at

least the jury is still out. Nonetheless, the case of deaf ASL-signers begins to muddy the waters. It is at least unclear whether a discrete format of representation is required. Hence, it is not obvious that language is special in its ability to induce categorical perception.

Nonetheless, it is still worth asking *how* gesture (whether combined with language or not) enhances performance on cognitive tasks. In other words, setting aside the issue of whether a discrete format of representation is required, what explains gesture's ability to boost cognitive performance? Does it involve noise reduction of representations?

Khatin-Zadeh et al. (2023) argue that gesture *schematizes* representations through abstraction of information. As the authors put it:

> [G]estural schematization promotes [generalization] through abstraction. This is a mechanism through which concrete knowledge is transformed into abstract knowledge. Gestural schematization discards irrelevant concrete details and allows individuals to focus their attention on a small set of essential elements. These elements are usually related to the global structure of a concept or…concepts. The global structure that is created through gestural schematization clearly depicts how spatial elements of a concept or… system of concepts are related to one another and how they interact with each other. This global or general structure is the product of a process through which concrete, unimportant, or irrelevant elements are deleted. The global structure is in fact a schema. The features of the global structure may be shared by a large number of concepts or systems of relations in a large number of contexts (2023, pg. 953).

As a toy example, consider learning the concept DOG. For the sake of simplicity, assume that a dog is a furry, four-legged animal that says "woof!". When you encounter a dog for the first time, you see its color, its size, the way it smells, that it wags its tail, etc. These features are not important to inferring category membership. However, you also see that it has fur, it's four-legged, and it says "woof!". The teacher points to the dog and says "This is a dog.". If gestures schematize representations through abstraction, then your representation of the dog will be

abstracted to include just those features relevant to category membership. In other words, the irrelevant features are deleted, and a global structure of the DOG concept is revealed.

Now, obviously merely pointing at the dog one time is not going to schematize the representation. In the Singer and Goldin-Meadow (2005) study above, the teacher iterated the lecture (with gestures) on mathematical equivalence multiple times. It is through this protracted training that gestures schematize. And importantly, this is not different from language's effects on cognition either. In the Lupyan et al. (2007) alien study, there was a training phase with repeated exposures and reinforcement. The effects of labeling and gesture are not one-shot.

If this is right, then gesture can augment representations in the same way that language does–through abstraction of task-irrelevant information. Although it is unclear whether gestural schematization requires a discrete representation, it is equally unclear whether language is uniquely capable of making representations more categorical. Hence, we can formulate more general hypotheses about whether a discrete representational format is required for abstraction of task-irrelevant information:

3. *Generalized Format Specific Thesis*: A necessary condition for representational abstraction is that two representations are combined, and (at least) one of which is a discrete representation.

4. *Generalized Format Agnostic Thesis*: It is not necessary for representational abstraction that when two representations are combined one of them is discrete.

Here, I have generalized away from gesture, and I used the "representational abstraction" to refer to the kind of abstraction involved in both gestural schematization and categorical perception, namely, the removal of task- and category-irrelevant information. In the next section, I will be showing that the *Generalized Format Specific Thesis* is false. Before moving on, I want to put forward a general description of *representational abstraction*–the kind that applies both to language and gesture in the cases of categorical perception and gestural schematization

respectively. This will serve as a functional description to assess whether the case of synesthesia I will be reviewing qualifies as representational abstraction.

Perception is messy and continuous. We encode far more information in perceptual representations than we need for many kinds of tasks (e.g., categorization, navigation, etc.). Language and gesture can help us because we can associate perceptual representations with them. This complicates things initially because it adds another representation into the problem solving space. However, after repeated exposure to these associations– after training with them– language and gesture bind to the task-relevant features of the target. In the cases of categorization and discrimination, labels and gestures become bound to the most diagnostic features of the target category. In the case of labels, this leads to more categorical representations; in the case of gestures, this leads to gestural schematization. I have argued that both of these processes involve reduction of task-irrelevant information through abstraction. In other words, both of these processes make representations simpler and less noisy. These simpler, less noisy representations are easier to link to lexical representations and they facilitate knowledge transfer. In the following sections, I will be extending this general description to the case of olfactory-color synesthesia, showing how synesthetic associations to color augment perceptual representations of smells, making them more categorical. If successful, this will show that discrete representations are not special in their ability to make representations more categorical. In other words, the *Generalized Format Specific Thesis* is false.

4. Olfactory-Color Synesthesia and Reduction of Odor Noise

4.1 Olfactory Categorization and Naming is Hard

One of the reasons the label superiority effect (i.e. categorical perception) and gestural schematization are so useful is because sometimes the world is really complex. Some objects in

the environment are perceptually complex–they have lots of properties and it's not always obvious which are relevant to category membership. Many categories overlap in features too. Some concepts, like mathematical equivalence, are really abstract and so do not have obvious perceptual extensions. Having mechanisms that facilitate noise reduction on highly abstract and perceptually complex representations allows us to better cope in the world. Indeed, without access to such mechanisms we might not be capable of solving some problems at all.

Olfactory identification and categorization are notoriously difficult for English speakers. A recent study by Majid and Burenhult (2014) found that English speakers overwhelmingly give source-based descriptions for odors in free naming smell tests, e.g., "This smells like Big Red gum." instead of naming the odor itself, e.g., "Cinnamon". However, different participants produced different source names for the same odor.[38]

At the same time, the neurobiology of olfaction reveals that the coding system is fundamentally different, and far more complex than, for example, vision. In particular, the olfactory system combinatorially codes odors with over 400 receptors. One of these receptors can respond to multiple different features of different odors, and one odorant can be coded by multiple different receptors (Malnic et al. 1999). Visual color perception (and color discrimination in particular) implicates only 3 different types of receptors, but olfactory processing implicates over 400 (Olender et al., 2012). And of course the number of combinations with a set of 400 is much larger than with a set of 3. Estimates of dimensions required for a specific number of molecules are provided by Young et al. (2014):

[38] See also Levinson and Majid (2014), San Roque et al. (2015), Cain (1979), Majid and Kruspe (2018), Olofsson and Wilson (2018), and Majid et al. (2018).

> Two odorous molecules and all their mixtures fill a one-dimensional [quality space]. Four molecules and their mixtures fill a two-dimensional [quality space]...To accommodate 100,000 odorous molecules and their mixtures, one would need a 17-dimensional space (pp.5-6).

Olfactory researchers Mamlouk and Martinez (2004) construct a perceptual space for olfactory perception based on 851 stimuli and 278 descriptions provided by Chee-Ruiter (2000). Using multidimensional scaling and principal component analysis, these researchers constructed a space with a lower bound of 32 dimensions and upper bound of 68 dimensions.

Another space was constructed by Castro et al. (2013) using principal component analysis and non-negative matrix factorization (a dimensionality reduction technique). The researchers constructed a 10 dimensional space that revealed discrete clustering of perceptual qualities. While this space certainly has fewer than 32 dimensions, Castro et al. used descriptors from Dravnieks' (1985) *Atlas of Odor Character Profiles.* Dravnieks used only 146 descriptors which is significantly less than those provided by Chee-Ruiter (2000) and used by Mamlouk and Martinez (2004). Moreover, setting aside the difference in the number of descriptors, a 10 dimensional space is still much higher than that of color visions'.[39]

The consensus is that olfaction is incredibly perceptually complex (see also Barwich 2019; Deroy 2023; Jraissati and Deroy 2021). This means that olfactory categorization and naming is hard. This makes olfaction a great candidate for categorical perception. Recent cross-linguistic analyses suggest that speakers of a language with abstract smell labels perform much better than speakers who lack such labels (Majid and Burenhult, 2014; Majid et al., 2018). One such study (Majid and Burenhult, 2014) compared performance between English and Jahai

[39] More discussion on the dimensions of olfactory perception space is provided by Meister (2015).

speakers. Jahai has abstract smells labels, English does not.[40] For example, "'*CNes*' is used for the smell of petrol, smoke, bat droppings and bat caves, some species of millipede, root of wild ginger, leaf of gingerwort, wood of wild mango, among other odor sources" (Majid and Burenhult 2014, pg.267). This word picks out a cluster of smells according to which qualities they are perceived to share.

On a free-naming smell identification test (modified B-SIT; see Doty et al., 1984), Jahai speakers produced abstract words the majority of the time, whereas English speakers primarily produced the source-based descriptions mentioned above. More important for present purposes, there was more agreement between Jahai speakers than between English speakers. The authors also report that English speakers struggled to name the smells, but the Jahai did not. It seems that if you speak an olfactory rich language, then you perform better on olfactory naming and categorization tasks. The label superiority effect rears its head. If you don't speak an olfactory rich language, then you will likely not perform so well on such tasks. Unless you are an olfactory-color synesthete.

4.2 Olfactory-Color Synesthesia and Superior Odor Naming

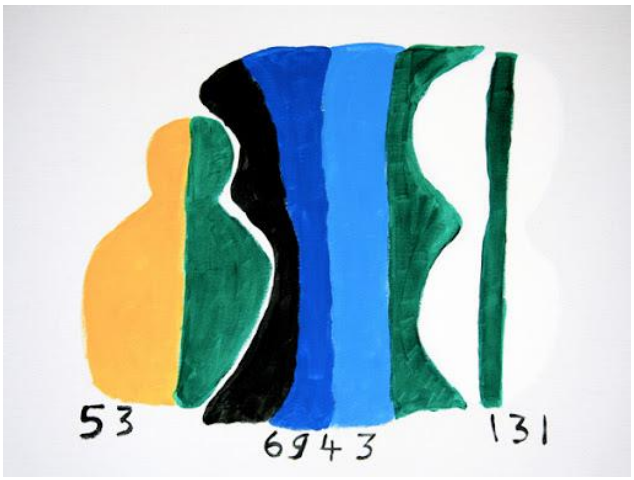
Synesthesia is often described as a condition in which stimulation in one sensory modality (e.g., vision) automatically causes stimulation in another modality (e.g., audition) (Baron-Cohen and Harrison 1997; Hubbard and Ramachandran 2005; Ramachandran and Hubbard 2001).[41] It is standard nomenclature to refer to the initial stimulus as the "inducer" and

---

[40] This is not quite accurate. "Musty" is an abstract smell label in English, but English speakers are not expected to become expert users of these labels, in the way that some other cultures are.

[41] This is actually not quite right, but will suffice for present purposes. A better analysis would be: stimulation in one sensory modality or cognitive faculty causes automatic stimulation (i) in another or the same sensory modality and/or (ii) in another or the same cognitive faculty. This is because there are cases of synesthesia where the inducing and/or the concurrent stimulation is cognitive. For example, consider time-unit synesthesia where weekdays or months are the inducers and colors and/or spatial configurations are the concurrents. Here, the inducers are not

the resultant stimulus as the "concurrent". For example, in grapheme-color synesthesia, the inducers are graphemes and the concurrents are colors.[42] Many synesthetes enjoy cognitive advantages seemingly attributable to their synesthesia. An extreme case of this is the mathematical savant, Daniel Tammet. Daniel Tammet has number-form-color synesthesia and Asperger's Syndrome (Baron-Cohen et al. 2007; Bor et al. 2008). He experiences numbers as colored and textured shapes (see picture 7). He recently won the prize of European Champion for memorizing the first 22,514 digits of Pi and is famous for his ability to carry out complex calculations in record time, e.g., multiplying 2 6-digit numbers in seconds. He reports that he performs these calculations by placing the two numer-forms next to one another. He then sees the negative space between them as the product (see picture 7).



PICTURE 7. A painting of how Daniel Tammet multiplies numbers. 53 times 131 is equivalent to the negative space between which is 6943. http://www.danieltammet.net/artwork.php

I am going to focus on a particular study done by Speed and Majid (2018) on olfactory-color synesthetes. In this kind of synesthesia, odors are the inducers and colors are the

perceptual stimuli-- one cannot perceive Thursday or February-- nonetheless these conceptual items cause sensory representations (in the case of color) and cognitive representations (in the case of spatial configurations).

[42] There are further distinctions with respect to the concurrents. For example, projectors experience the concurrents as part of the inducing stimuli (e.g., the grapheme is literally green), whereas associators experience the concurrent "in the mind's eye".

concurrents. That is, olfactory-color synesthetes have particular colors automatically and consistently elicited by particular smells. Over a period of two days, Speed and Majid conducted various tests with controls and synesthetes including: odor naming, odor discrimination, color discrimination, and odor threshold. For present purposes, I will focus just on performance in the odor naming and odor discrimination tasks. Naming tasks ask participants to identify the stimulus they are presented with. For example, in this study, participants were presented with odors and asked "What smell is this?". Discrimination tasks ask participants whether two stimuli are different or the same. For example, if presented with three smell sticks (two with the same scent and one different), I will have successfully discriminated between them if I choose the one that is different.

The results of the study showed that synesthetes significantly outperformed controls on both odor naming and discrimination. This result is in line with previous studies reporting superior naming and discrimination performance among synesthetes (Chiou and Rich 2014; Deroy and Spence 2013; Kadosh et al. 2009; Mroczko-Wąsowicz and Werning 2012; Radvansky et al. 2011; Rich and Karstoft 2013; Spector and Maurer 2009; Ward et al. 2010; Watson et al. 2014; Witthoft and Winawer 2013). Why would color representations make a difference to naming and discrimination performance? Speed and Majid provide a nice explanation. Here's what they say:

> Connections between odor concepts and other sensory areas of the brain, such as vision, can similarly improve odor naming ability…We propose synaesthetic associations to odors act as additional semantic features, increasing the semantic richness of the odor concept, and thereby facilitating odor naming (2018, pg.478).

For Speed and Majid, color representations become a part of odor concepts. This, they claim, *enriches* odor concepts making them more differentiated and easier to link to lexical representations (pg. 478). The former idea makes discrimination easier: if odor representations are more differentiated, then they are more categorical. They are more categorical because the between-category differences are more salient than within category differences. The latter idea makes naming easier: if odor concepts are more categorical, then inferring category membership is easier. On this view, odor representations become more complex, but that complexity, it is supposed, issues in cognitive advantages.

I agree with Speed and Majid that color representations become a part of odor concepts. However, we should be clear about what it means to "enrich" those concepts. The general description of representational abstraction that I provided above says that the addition of, for example, a label, initially complicates the problem solving space. Over time though, the label or the gesture becomes bound to the category or task relevant features of the target, and noise is reduced through abstraction. This leaves the subject with a simplified, less noisy representation that is easier to link to lexical items and which facilitates knowledge transfer and generalization. I think given this, the best explanation for the synesthetes' superior performance is that color representations reduce category irrelevant information in odor representations. This makes them more differentiated and easier to link to lexical representations. If this is right, then synesthetic concurrents enhance cognitive performance in the same way that language and gesture does.[43]

[43] It's also worth saying that Speed and Majid (2018) note that connections between odor concepts and other sensory areas of the brain improve naming. The reason for this is that combining information from multiple modalities enhances representation of the target category. This enhancement involves a reduction of noise. Perceptual cue integration models (Ernst and Banks 2002) posit that cues from multiple modalities get optimally combined by averaging the sources together weighted by the certainty that accompanies each source of information. The result is an abstract representation of the category (see also Reiger and Xu 2017 for an application of this model to categorical perception).

Moreover, since olfactory and color representations are paradigmatically non-discrete, this shows that the *Generalized Format Specific Thesis* is false.

4.3 Neural Evidence of Common Mechanism

To strengthen the inductive support of my proposal, it is worth considering the neural mechanisms that might subserve this binding and subsequent reduction of noise across synesthesia, gesture, and the label superiority effect. Taking my cue from the quote by Speed and Majid above, what neural mechanism(s) could be responsible for crossmodal bindings (bindings of representations from different modalities)?

Initially introduced by Damasio (1989), the convergence zone framework posits a hierarchical binding process. In first-order convergence zones, early sensorimotor representations are bound together. In second-order zones, representations from the first-order zones are bound. The hierarchical processing continues, issuing in increasingly abstract representations. Importantly, the binding process involves noise reduction through similarity computations and compression algorithms. This framework has been adapted by various other researchers, including Barsalou's Perceptual Symbol Systems (1999), Lambon Ralph et al.'s (2010) Hub and Spoke Model, Reilly et al.'s (2016) Dynamic Multilevel Reactivation Framework, Binder and Desai's (2011) Embodied Abstraction framework, and Dove's (2022) Elastic Mind and LENS theory.

One finding that remains constant across all these iterations of Damasio's original framework is that there are regions of the brain that take in inputs from multiple modalities, bind them together, and output a coherent, more abstract representation of the target. The anterior temporal lobe (ATL) has been a region of interest since the Hub and Spoke model claimed that it was *the* central convergence hub (Lambon Ralph et al. 2010). Another region that has gained

interest is the angular gyrus in the inferior parietal lobule (Bonner et al. 2013).[44] Through fMRI and DTI (diffuser tension imaging) in a lexical decision task, Bonner et al. found that:

> [T]he angular gyrus is consistently activated by concepts with different modality-specific associations and… the angular gyrus has white matter connectivity to regions of modality-specific feature representation (pg. 182).

The idea is that concepts rely on heteromodal convergence zones and modality specific representations in sensory, motor, and effective areas. Bonner et al. think that they have good evidence that the angular gyrus is at least one of these heteromodal convergence zones. Indeed, Davis and Yee (2019) have argued that these areas (the ATL and the angular gyrus) constitute the two convergence hubs that underwrite human semantic memory. For my part, there is evidence that both the ATL and angular gyrus are implicated in the label superiority effect, gestural schematization, and synesthesia. If these areas are all implicated in these phenomena, and these areas are responsible for binding and abstracting from multimodal inputs, then we have good reason to think that these phenomena are best explained by binding and abstraction in heteromodal convergence zones.

There is overwhelming evidence that both the ATL and the angular gyrus are implicated in lexical decision tasks (Davis and Yee 2019; Seghier 2013; Zhang et al. 2024). Simanova et al. (2016) review a wealth of studies indicating convergence zone activation (including ATL and angular gyrus) during multimodal object recognition (see also Man et al. 2013). Most notably, however, Perry and Lupyan (2014) hypothesized that direct cathodal stimulation of Wernicke's

---

[44] Geschwind (1972) might be the earliest work suggesting the angular gyrus is a heteromodal convergence zone. However, there are other regions of interest on this front as well. Indeed, it is very likely the case that there are multiple higher-level, heteromodal and supramodal convergence zones (Dove 2022). See also Binder and Desai (2011).

area (in the ATL, and bordering the angular gyrus), would disrupt categorization. The authors report that if this hypothesis is confirmed, then this is positive evidence that "language is involved in the ability to form object representations that emphasize task-relevant dimensions" (pg. 67). Indeed, this is precisely what they find. Hence, convergence zone mechanisms are implicated in the label superiority effect and categorical perception.

There is also good evidence that the angular gyrus is implicated in gestural schematization. Wakefield et al. (2019) used fMRI to measure activations in children solving mathematical equivalence problems. Children who gestured significantly outperformed the children who did not. Moreover, Wakefield et al. report that children who gestured had increased angular gyrus activations. The authors note that:

> [T]he AG [angular gyrus] is involved in mapping math problems to visuo-spatial referents. The fact that the AG was more heavily recruited by children in the speech + gesture group than by children in the speech-alone group might therefore reflect gesture's capacity to provide a spatial framework for mathematical equivalence problems (pg. 2351).

Hence, we have good reasons for thinking that gestural schematization implicates convergence zones, and in particular the angular gyrus.

Finally, there is evidence that synesthesia implicates convergence zones too. Rouw et al. (2011) review various behavioral and imaging studies on various kinds of synesthesia. One of the most consistent findings is the activation of the angular gyrus. As they summarize the data: "All locations of activation in inferior parietal lobule, however, are best summarized as either near the intraparietal sulcus or in the angular gyrus" (pg. 226). They go on to say that this finding is not surprising given that synesthesia involves bindings between modalities, and the angular gyrus is known for its contributions to binding more generally. Additionally, Chiou and Rich

(2014) review evidence that the ATL is implicated in synesthetic color representation, specifically drawing upon the Hub and Spoke model of conceptual processing (Chiou et al. 2014; van Leeuwen et al. 2010).

In sum, we have good reasons for thinking that the synesthetic performance advantages in Speed and Majid's study are explainable by the addition to color information to odor representations. This addition does indeed enrich these representations, making them more differentiated, and easier to link to lexical representations. However, the best explanation for how these representations become more differentiated involves a reduction of information through binding and abstraction mechanisms found in convergence zones, particularly the ATL and the angular gyrus. If this is right, then synesthetic performance advantages, the label superiority effect, and gestural schematization are explained by a common mechanism. It also follows that language is not special in its ability to make representations more categorical because color associations can make olfactory representations more categorical. Hence, the *Generalized Format Specific Thesis* is false. This view fits nicely with a more general trend in semantic memory research as viewing concept formation and deployment as multi-level and multi-modal (Dove 2022). In the final section, I will address three powerful objections.

5. Objections

Lupyan and Thompson-Schill (2013) did an experiment testing whether non-linguistic cues to categories could induce categorical perception. They tested subjects by exposing them to a category label or a highly associatively linked non-linguistic cue before asking them to select the target category exemplar from an array. For example, in being exposed to either the label "dog" or the sound of a dog barking, the subject must select the dog image from the array. The authors found that there was only a reaction time and accuracy advantage for the label condition,

not the non-linguistic cue condition. This study shows that labels are special in activating more categorical representations; non-linguistic perceptual cues do not induce categorical perception. Hence, it is false that synesthetic concurrents induce categorical perception.

In responding to this objection, it is not sufficient to cite that the synesthetic concurrents are intimately associatively linked to the category. This won't do because surely dog barks are associatively linked to the dog category. However, if that's true, then why do dog barks not induce categorical representations of the dog category? When you hear a dog bark, it conveys specific information about particular dogs– high pitched barks correlates with small dogs, low pitched barks correlates with big dogs. Dog barks are not sufficiently distanced from particular category exemplars to induce a more abstract, categorical representation. The same is true for color experiences though. When you experience redness, you get specific information about hue, brightness, and saturation. The difference with synesthesia, however, is that olfactory-color synesthetes are not using color concurrents to identify colors, they are using them to identify smells. There is nothing about experiencing a particular shade of red that would tell you anything about a smell category–olfactory-color associations are arbitrary, just like category labels and categories. There is nothing about the word "dog" that makes it an especially good cue to the dog category, any other word would have done the job. This is a benefit because unlike dog barks, "dog" is sufficiently distanced from category exemplars. Likewise, there is nothing about a shade of red that makes it an especially good cue to the, say, rosemary category. But once these arbitrary associations (word–category, color–category) are in place, they abstract over particulars of the categories and stand-in for the category as a whole. In essence, the color concurrents, despite actually being non-linguistic perceptual cues, are functioning like discrete representations, at least in one respect.

Consider another objection: it's false that the synesthetic performance advantages show that discrete representations are not special in their ability to make representations more categorical because it's possible that language is making an online contribution to the synesthetes' performance. The idea might be that colors are highly lexicalized in English so when the synesthetes experience a color concurrent, they use the color category label to link to olfactory labels. In this case, it's not true that the color representations are being bound to olfactory representations and abstracted. Instead, the subjects simply use the color category labels as guides to the olfactory category labels. In other words, odor-color associations are mediated by language.

In response, it's important to see that this view predicts that verbal interference should disrupt these associations. Recall that in the Winawer et al. (2007) study on Russian versus English categorization of blues, the authors found that verbal interference reduced the Russian advantage. This suggests that labels are making the representations more categorical online. Similarly, if labels are doing the work online in odor-color synesthesia, then verbal interference should disrupt performance.

Speed et al. (2023) tested this exact prediction, but with non-synesthetes. The authors first deployd a norming study gathering data on Dutch speakers' odor-color associations. For example, the smell of coconut is associated with white, banana with yellow, lemon with yellow, etc. This norming study was used to test for odor-color matching in the main experiments. In the main experiments, Speed et al. used verbal interference during an odor-color matching task and an odor naming task. Between the two experiments, they found no effect of verbal interference. However, they found that there was higher color matching for correctly named odors and more familiar odors. The authors report that these results indicate that language does not make online

contributions to odor-color associations. For my part, I think this is good evidence that the above objection fails. Language is not mediating synesthetic performance. In fact, there is positive evidence that the color associations are actually mediating performance.

Finally, I will consider one final objection. This objection agrees that color associations are making odor representations more categorical. It is also agrees that language is not making an online contribution. However, the objector says that the only reason color representations could make odor representations more categorical is that they are discretized by language in long-term memory. In other words, discrete representations are still special because color representations are emulating discrete representations in making odor representations more categorical.

In response, this view entails that color representations are stored as categorical or discrete and retrieved as such. Hence, this view predicts that verbal interference on color categorization and discrimination tasks should have no effect. However, this prediction is false. As already discussed, Winawer et al. (2007) showed that the Russian speakers' advantage in color categorization and discrimination vanished during verbal interference. This suggests that language is making an online contribution to color representations–namely, making them more categorical. Hence, it is false that color representations are stored as categorical, discrete representations in long-term memory. Therefore, I submit that discrete representations are not special in their ability to make representations more categorical– the *Generalized Format Specific Thesis* is false. Olfactory-color synesthetes' cognitive advantage in odor naming is best explained by non-discretized color representations binding with olfactory representations and becoming more categorical through abstraction processes probably in convergence zones.

Before ending this section, I want to draw out a tension that the reader may have picked up on. Speed et al. (2023) found no effect of verbal interference on odor-color associations and

naming which suggests that language is not making an online contribution. However, Winawer et al. (2007) did find an effect of verbal interference on color categorization and discrimination which suggests that language does make an online contribution. This might seem initially puzzling: why would language only sometimes make an online contribution? The difference, I suggest, is that in the Winawer et al. study, subjects did not have access to other kinds of representation. Verbal interference prohibited access to lexical representations, so the subjects had to rely solely on perceptual representations to discriminate and infer category membership. Whereas in the Speed et al. (2023) study, the subjects were explicitly told to associate the odors with colors. This task essentially involves recruiting representations in another modality. Further, verbal interference prohibited online linguistic effects, so the subjects relied solely on the olfactory and color representations.

I submit then that the *Generalized Format Specific Thesis* is false. It is not a necessary condition on representational abstraction to combine two representations, (at least) one of which is discrete. Combining two analog representations is sufficient for representational abstraction. Indeed, I have argued that there is a common mechanism that explains representational abstraction across categorical perception, gestural schematization, and synesthesia: the reduction of noise through binding and abstraction in convergence zones.

6. Conclusion

In this paper, I have argued that discrete representations are not unique in their ability to make representations more categorical. That is, the *Generalized Format Specific Thesis* is false. I showed how in olfactory-color synesthesia color associations make olfactory representations more categorical, and this explains the synesthetes' superior olfactory categorization and discrimination performance. Along the way, I surveyed the literature on the cognitive advantages

of analog gesture. Gesture is an interesting test case for the *Generalized Format Specific Thesis* because one plausible explanation for gesture's cognitive advantages is that gestures schematize perceptual representations by binding and abstracting them to be less noisy. However, the jury is still out on whether gestural schematization requires a discrete representation to be bound, as the case of ASL-signers showed. At this point, it is unclear whether ASL is analog or discrete. Hence, gesture cannot be used to decide whether language (or a discrete representational format more generally) is necessary for making representations more categorical. However, the case of olfactory-color synesthesia does decide this question. As I showed in responding to objections, language does not mediate odor-color associations online, nor is it true that color representations are stored in a discrete format, emulating language.

My arguments in this paper provide a plausible common mechanism for representational abstraction induced by language, gesture, and synesthesia. Indeed, I showed that so-called "convergence zones", plausibly located in the angular gyrus and anterior temporal lobes, are implicated in the label superiority effect, gestural schematization, and synesthesia. This finding is actually unsurprising given the proposed function of convergence zones: to bind and abstract across multimodal inputs.

Future work will hopefully decide whether ASL counts as a discrete system or not. In either case, further work on gesture in deaf signers is warranted. An upshot of the view proposed here is that non-linguistic creatures can benefit from more categorical perception because language is not required. Comparative psychological work investigating this prediction would be useful. Additionally, further neuroimaging work investigating my hypothesis about a common mechanism, involving convergence zones, for categorical perception is welcome.

REFERENCES

Baron-Cohen, S., & Harrison, J. (1997). Synaesthesia: a review of psychological theories. *Synaesthesia: Classic and Contemporary Readings, Oxford: Blackwell*, 109-22.

Baron-Cohen, S., Bor, D., Billington, J., Asher, J., Wheelwright, S., & Ashwin, C. (2007). Savant memory in a man with colour form-number synaesthesia and asperger. *Journal of consciousness studies*, *14*(9-10), 237-251.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(4), 577-660.

Barwich, A. S. (2019). A critique of olfactory objects. *Frontiers in Psychology*, *10*, 453626.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, *15*(11), 527-536.

Bonner, M. F., Peelle, J. E., Cook, P. A., & Grossman, M. (2013). Heteromodal conceptual processing in the angular gyrus. *Neuroimage*, *71*, 175-186.

Bor, D., Billington, J., & Baron-Cohen, S. (2008). Savant memory for digits in a case of synaesthesia and Asperger syndrome is related to hyperactivity in the lateral prefrontal cortex. *Neurocase*, *13*(5-6), 311-319.

Cain, W. S. (1979). To know with the nose: keys to odor identification. *Science*, *203*(4379), 467-470.

Castro, J. B., Ramanathan, A., & Chennubhotla, C. S. (2013). Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. *PloS one*, *8*(9), e73289.

Chee-Ruiter, C. W. (2000). *The biological sense of smell: olfactory search behavior and a metabolic view for olfactory perception*. California Institute of Technology.

Chiou, R., & Rich, A. N. (2014). The role of conceptual knowledge in understanding synaesthesia: Evaluating contemporary findings from a "hub-and-spokes" perspective. *Frontiers in Psychology*, *5*, 64075.

Chiou, R., Sowman, P. F., Etchell, A. C., & Rich, A. N. (2014). A conceptual lemon: Theta burst stimulation to the left anterior temporal lobe untangles object representation and its canonical color. *Journal of cognitive neuroscience*, *26*(5), 1066-1074.

Clark, A. (2006). Material symbols. *Philosophical psychology*, *19*(3), 291-307.

Clark, A. (2012). Magic words: How language augments human computation. In *Language and meaning in cognitive science* (pp. 21-39). Routledge.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural computation*, *1*(1), 123-132.

Davis, C. P., & Yee, E. (2019). Features, labels, space, and time: Factors supporting taxonomic relationships in the anterior temporal lobe and thematic relationships in the angular gyrus.

Language, Cognition and Neuroscience, *34*(10), 1347-1357.

Delgado, B., Gómez, J. C., & Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: Experimental evidence. *Journal of experimental child psychology*, *110*(3), 299-312.

Deroy, O. (2023). Olfactory abstraction: a communicative and metacognitive account. *Philosophical Transactions of the Royal Society B*, *378*(1870), 20210369.

Deroy, O., & Spence, C. (2013). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1240-1253.

Doty, R. L., Shaman, P., & Dann, M. (1984). Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiology & behavior*, *32*(3), 489-502.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412-431.

Dove, G. (2020). More than a scaffold: Language is a neuroenhancement. *Cognitive neuropsychology*, *37*(5-6), 288-311.

Dove, G. (2022). *Abstract concepts and the embodied mind: Rethinking grounded cognition*. Oxford University Press.

Dravnieks, A. (1985). *Atlas of odor character profiles*.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

Gentry, H. (Forthcoming). Constructing Embodied Emotion with Language: Moebius Syndrome and Face-Based Emotion Recognition Revisited. *Australasian Journal of Philosophy.*

Geschwind, N. (1972). Language and the brain. *Scientific American*, *226*(4), 76-83.

Goldin-Meadow, S. (2015). From action to abstraction: Gesture as a mechanism of change. *Developmental review*, *38*, 167-184.

Goldin-Meadow, S., Shield, A., Lenzen, D., Herzig, M., & Padden, C. (2012). The gestures ASL signers use tell us when they are ready to learn math. *Cognition*, *123*(3), 448-453.

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and brain sciences*, *40*, e46.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27-43.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69-78.

Holmes, K. J., & Regier, T. (2017). Categorical perception beyond the basic level: The case of warm and cool colors. *Cognitive science*, *41*(4), 1135-1147.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological bulletin, 137*(2), 297.

Hubbard, E. M., & Ramachandran, V. S. (2005). Neurocognitive mechanisms of synesthesia. *Neuron*, *48*(3), 509-520.

Jraissati, Y., & Deroy, O. (2021). Categorizing smells: A localist approach. *Cognitive Science*, *45*(1), e12930.

Kadosh, R. C., Henik, A., & Walsh, V. (2009). Synaesthesia: learned or lost?. *Developmental Science*, *12*(3), 484-491.

Khatin-Zadeh, O. (2022). How does representational transformation enhance mathematical thinking?. *Axiomathes*, *32*(Suppl 2), 283-292.

Khatin-Zadeh, O., Farina, M., Yazdani-Fazlabadi, B., Hu, J., Trumpower, D., Marmolejo-Ramos, F., & Farsani, D. (2023). The roles of gestural and symbolic schematizations in inhibition as a component of executive functions. *Integrative Psychological and Behavioral Science*, *57*(3), 950-959.

Lambon Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, *107*(6), 2717-2722.

Levinson, S. C., & Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, *29*(4), 407-427.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in psychology*, *3*, 54.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological science*, *18*(12), 1077-1083.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170.

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1319-1337.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in cognitive sciences*, *24*(11), 930-944.

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), 266-270.

Majid, A., & Kruspe, N. (2018). Hunter-gatherer olfaction is special. *Current Biology*, *28*(3), 409-413.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'grady, L., ... & Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369-11376.

Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell*, *96*(5), 713-723.

Mamlouk, A. M., & Martinetz, T. (2004). On the dimensions of the olfactory perception space. *Neurocomputing*, *58*, 1019-1025.

Man, K., Kaplan, J., Damasio, H., & Damasio, A. (2013). Neural convergence and divergence in the mammalian cerebral cortex: from experimental neuroanatomy to functional neuroimaging. *Journal of Comparative Neurology*, *521*(18), 4097-4111.

Meister, M. (2015). On the dimensionality of odor space. *Elife*, *4*, e07865.

Mroczko-Wąsowicz, A., & Werning, M. (2012). Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Frontiers in Psychology*, *3*, 279.

Olender, T., Waszak, S. M., Viavant, M., Khen, M., Ben-Asher, E., Reyes, A., ... & Lancet, D. (2012). Personal receptor repertoires: olfaction as a model. *BMC genomics*, *13*, 1-16.

Olofsson, J. K., & Wilson, D. A. (2018). Human olfaction: It takes two villages. *Current Biology*, *28*(3), R108-R110.

Pérez-Gay Juárez, F., Thériault, C., Gregory, M., Rivas, D., Sabri, H., & Harnad, S. (2017). How and why does category learning cause categorical perception?. *International journal of comparative psychology*, *30*.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66-72.

Radvansky, G. A., Gibson, B. S., & McNerney, M. (2011). Synesthesia and memory: color congruency, von Restorff, and false memory effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 219.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of consciousness studies*, *8*(12), 3-34.

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(6), e1440.

Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: the dynamic multilevel reactivation framework. *Psychonomic bulletin & review*, *23*, 1002-1014.

Rich, A. N., & Karstoft, K. I. (2013). Exploring the benefit of synaesthetic colours: Testing for “pop-out” in individuals with grapheme–colour synaesthesia. *Cognitive neuropsychology*, *30*(2), 110-125.

Rouw, R., Scholte, H. S., & Colizoli, O. (2011). Brain areas involved in synaesthesia: a review. *Journal of neuropsychology*, *5*(2), 214-242.

Russell, J. A., & Widen, S. C. (2002). A label superiority effect in children's categorization of facial expressions. *Social Development, 11*(1), 30-52.

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., ... & Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive linguistics*, *26*(1), 31-60.

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist*, *19*(1), 43-61.

Simanova, I., Francken, J. C., de Lange, F. P., & Bekkering, H. (2016). Linguistic priors shape categorical perception. *Language, Cognition and Neuroscience*, *31*(1), 159-165.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological science*, *16*(2), 85-89.

So, W. C., Ching, T. H. W., Lim, P. E., Cheng, X., & Ip, K. Y. (2014). Producing gestures facilitates route learning. *PloS one*, *9*(11), e112543.

Spector, F., & Maurer, D. (2009). Synesthesia: A New Approach to Understanding the Development of Perception. *Developmental Psychology*, *45*(1), 175-189.

Speed, L. J., & Majid, A. (2018). Superior olfactory language and cognition in odor-color synaesthesia. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(3), 468.

Speed, L. J., de Valk, J., Croijmans, I., Huisman, J. L., & Majid, A. (2023). Odor-Color Associations Are Not Mediated by Concurrent Verbalization. *Cognitive science*, *47*(4), e13266.

Van Leeuwen, T. M., Petersson, K. M., & Hagoort, P. (2010). Synaesthetic colour in the brain: beyond colour areas. A functional magnetic resonance imaging study of synaesthetes and matched controls. *PloS one*, *5*(8), e12074.

Wakefield, E. M., Congdon, E. L., Novack, M. A., Goldin-Meadow, S., & James, K. H. (2019). Learning math by hand: The neural effects of gesture-based instruction in 8-year-old children. *Attention, Perception, & Psychophysics*, *81*, 2343-2353.

Ward, J., Jonas, C., Dienes, Z., & Seth, A. (2010). Grapheme-colour synaesthesia improves detection of embedded shapes, but without pre-attentive ‘pop-out’of synaesthetic colour. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 1021-1026.

Watson, M. R., Akins, K. A., Spiker, C., Crawford, L., & Enns, J. T. (2014). Synesthesia and learning: a critical review and novel theory. *Frontiers in human neuroscience*, *8*, 98.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the national academy of sciences*, *104*(19), 7780-7785.

Witthoft, N., & Winawer, J. (2013). Learning, memory, and synesthesia. *Psychological science*, *24*(3), 258-265.

Young, B. D., Keller, A., & Rosenthal, D. (2014). Quality-space theory in olfaction. *Frontiers in Psychology*, *5*, 68136.

Zhang, Y., Wu, W., Mirman, D., & Hoffman, P. (2024). Representation of event and object concepts in ventral anterior temporal lobe and angular gyrus. *Cerebral Cortex*, *34*(2), bhad519.

Chapter 3: The Constructive Episodic-Semantic Memory System

1. Introduction

Originally formulated by Tulving (1972), the distinction between episodic and semantic memory has been a mainstay of memory research–leading to an understanding of these memory systems as constituting two distinct natural kinds. Episodic memory is thought to consist in autobiographical recall for specific events with content specifying "what, where, and when" and a mental time travel phenomenology (sometimes referred to as "autonoesis"). Semantic memory, on the other hand, is thought to consist in recall for depersonalized facts and abstract conceptual knowledge.

Many researchers now think that episodic memory is constructive (Aronowitz, 2019; de Brigard, 2014; Michaelian, 2011; 2016; 2022). The thought is that episodic memory is *not* archival, storing the entirety of the details of a token event. Instead, memory "traces" are stored, and through probabilistic mechanisms the event details are constructed at retrieval.

In this paper, I will be in agreement with some new memory research casting doubt on the claim that episodic and semantic memory are two distinct natural kinds (de Brigard, Umanth, and Irish, 2022; Donaldson, 1996; Dunn, 2004; Gentry and Buckner, forthcoming; Irish, 2020; McClelland, McNaughton, & O'Reilly, 1995; McKoon & Ratcliff, 1986; Robins, 2023; Rubin, 2022; Tanguay et al., 2023; Vatansever et al., 2021; Wixted, 2007).[45] I will be reviewing some neuroscientific evidence indicating that the episodic and semantic memory systems grade into one another along a continuum (at multiple levels of description). I will ultimately argue that if (i) episodic and semantic memory mechanistically grade into one another, and (ii) episodic

[45] There are also views expressing a skepticism about a categorical distinction, without explicitly committing to a continuum view (e.g., Cermak et al., 1978; Duff et al., 2020; Graham et al., 2000; Greenberg and Verfaellie, 2010; Squire et al., 1984).

memory is constructive, then we have good reasons to think that semantic memory is constructive too. The result is a singular, constructive episodic-semantic memory system.

I do *not* aim to show that, as a matter of fact, episodic and semantic memory fail to be categorically distinguishable. By critically engaging with the most popular arguments for drawing such distinctions, I aim to show that (at multiple levels of description) these systems are continuous with each other– in the sense of sharing encoding, storage, and retrieval properties. In addition to arguing for the overall similarity between these two systems, I will mount direct evidence that semantic memory is constructive, just like episodic memory.

Here's the roadmap for this paper. In section 1, I will recount arguments against a natural kind distinction between episodic and semantic memory. Here, I will conclude that the strongest argument shows that these memory systems instantiate the metaphysical phenomenon of transitional gradation at the level of mechanism (Buckner, 2016; Gentry and Buckner, forthcoming). In section 2, I will review evidence that episodic recollection is a constructive process. I will then present the main argument of this paper in section 3. That argument concludes that we have good reasons to think that semantic memory is constructive. Section 4 will conclude with a proof of concept for the view on offer–the semantic pointer architecture.

2. **Arguments Against a Natural Kind Distinction**[46]

2.1 Content-Based Criterion

The simplest criterion for cleaving episodic from semantic memory is a content-based criterion. Tulving (1972) thought that the representational content –"the nature of the stored information"–imputed to these mnemonic representations was sufficient to categorically distinguish them. Episodic memories were thus memories of events as perceived in one's own

[46] For a fuller treatment of these issues, see Gentry and Buckner (forthcoming). For the sake of space, I am fairly brief in my exposition.

subjective experience, subjectively contextualized by their "temporal-spatial relation to other experienced events" (Tulving 1972, pg. 388). This criterion led to the mantra that episodic memories encoded "what-where-when" content (Clayton & Dickinson, 1998; Russell et al., 2011). For example, suppose S recalls her 7th birthday party having a bounce house in Aurora, Illinois in mid-September 2001. S episodically recalls this event.

Semantic memories, on the other hand, were thought to consist in mere recall of abstract facts–depersonalized and decontextualized. The content of these memories lack "what-where-when" information. For example, S recalls that Abraham Lincoln was the 16th president of the United States, but fails to recall where she learned this, how she learned it, and when she learned it.

This content-based criterion fails to mark out a natural kind distinction between these memory systems. Semantic memory is flexible enough to represent the "what-where-when" content even in relation to other events. For example, the semantic memory 'that Christopher Columbus sailed the ocean blue in 1492' contains information about what happened, where it happened, and when it happened. Further, semantic memories could even be about autobiographical events, as when an adult cannot episodically recall a childhood event, but can tell you what happened, when it happened, and where it happened.

Indeed, such mnemonic contents are precisely what we find in cases of "repisodic" memories (Nessier, 1981; Renoult et al., 2012; Rubin & Umanath, 2015). Nessier (1981) described these memories as mnemonic representations for oft repeated events. For example, remembering a traditional holiday family dinner that occurs every year. Here, subjects extract themes or statistical regularities in repeated exposures to events and recollect those themes. However, it's not the case that they are recollecting any one particular event. This is evidenced

by the fact that subjects frequently get the details wrong. Consider what Nessier says about the former counsel to President Nixon who testified in the Watergate scandal, John Dean's memory:

> Such memories might be called repisodic rather than episodic: what seems to be an episode actually represents a repetition. Dean remembers the million-dollar remark because Nixon made it so often; he recalls the "cancer" metaphor because he first planned it and then repeated it; he remembers his March 21 lecture to the President because he planned it, then presented it, and then no doubt went over it again and again in his own mind. What he says about these "'repisodes" is essentially correct, even though it is not literally faithful to any one occasion. He is not remembering the "gist" of a single episode by itself, but the common characteristics of a whole series of events (1981, pg. 20).

You might think that repisodic memories are a small subset of memories and as such, don't pose much of a threat to a content-based criterion. You might think this because we are not repeatedly exposed to the same event for every event we experience. However, Walker and Stickgold (2010) argue that episodic memories are consistently replayed during REM sleep– a process they call "memory integration". Replay of memories overnight results in the extraction of themes and statistical regularities that are paradigmatic of repisodic memories.

2.2 Phenomenological Criterion

With the failure of the content criterion, memory theorists became increasingly reliant on phenomenology. So-called "autonoesis" (sometimes referred to as the feeling of ownership, mental time travel, or reliving) was proposed alongside the content-based criterion. The idea was that episodic memories essentially involve a particular phenomenology. So while it is true that semantic memories can plausibly represent 'what-where-when' content, they will fail to produce the required phenomenology.

However, this phenomenological criterion also fails. Not only do repisodic memories exhibit episodic-like contents, but they also exhibit episodic-like phenomenology. That is, subjects who repisodically remember have the feeling of reliving, even though they are not targeting a specific event, but a collection of repeated events (Nessier, 1981; Renoult et al., 2012).

The fact that repisodic memories can exhibit episodic phenomenology inspired the signal detection theory (Wixted and Stretch, 2004). The idea is that the sense of reliving is generated by a signal computed on the results of recollection. When the continuous dimensions of recollection exceed a threshold, the signal is activated and the subject feels the sense of the reliving. In other words, whether a subject reports episodic or semantic recollection depends on the richness of the content recollected. On this view, there aren't two different memory systems (episodic and semantic). Indeed, this view provides an error theory for why we think there are two different memory systems. The error theory holds that the phenomenology of recollection seems to be evidence of a sharp boundary between two kinds of memory, but in fact there is no such boundary.[47]

As a brief illustration, consider the case of patient RB. RB suffered from a traumatic head injury resulting in anterograde and retrograde amnesia. After recovering from his amnesias, he reported experiencing autobiographical memories in rich contextual detail, but not feeling as if they were his–as if they happened to him. Gentry (2023) argues that the reason RB lacks this feeling of ownership is *not* because he can't episodically recollect, but due to a signal detection deficit. RB's memories feel as if they didn't happen to him because the signal computed on the

[47] Another option is to ditch the content criterion and go for a full-throated phenomenological distinction (see Klein 2016). However, if you opt for this view, you run the risk of making the distinction between episodic and semantic memory empirically intractable in non-human animal research (see Gentry and Buckner, forthcoming).

results of recollection fail to generate the associated phenomenology. This explanation, argues Gentry, actually fits within a larger context of dissociative and self-related disorders (e.g., dissociative PTSD and depersonalization). The upshot of this argument is that episodic phenomenology comes apart from episodic content recall, indeed it is not necessary. Thus, a phenomenological criterion fails to cleave episodic and semantic memory.

2.3 Semanticization and Composition: Surface-Level Transitional Gradation

Another argument against a natural kind distinction comes from Sara Aronowitz's (forthcoming) work on semanticization. Semanticization involves migration from episodically encoded memories to semantically encoded memories.[48] More precisely:

> Information about past environments is initially stored and retrieved in an episodic form, with features such as concrete detail, imagery, and connection with other temporally related content, but over time and exposure, this information shifts away from these episodic features and is consolidated in an increasingly semantic form characterized by gist, abstraction, and connection with other semantically related content (Aronowitz forthcoming, pg.5).[49]

Aronowitz argues that if either the complementary learning systems theory or the navigational theory of memory is true, then semanticization follows. If semanticization occurs, then there is no natural kind distinction between episodic and semantic memory. This is because semanticization entails a proliferation of transitional forms of representation between episodic

---

[48] This process can be described at two different levels of analysis: the neural and the cognitive. Neural semanticization is when information about the past is initially encoded in neural systems tied to episodic storage and retrieval (MTL), but this information shifts from MTL to neocortical areas tied to semantic storage and retrieval. Cognitive semanticization, on the other hand, is when information about the past is initially stored and retrieved in episodic form (e.g., what-where-when content, concrete details, and imagery), but this information shifts to a semantic form (e.g., gist, abstraction).

[49] Semanticization is precisely the explanation for John Dean's memory recounted in Nessier (1981) and reviewed above. With repeated exposure, the specifics of repeated events get lost, but a more gist based, schema evolves.

and semantic forms. That is, the process of migrating from episodic to semantic is continuous, not dichotomous. If this is right, then any attempt to draw a categorical distinction will be arbitrary.

The complementary learning systems theory posits two different memory systems: the hippocampal (episodic) system and the neocortical (semantic) system, with a consolidation process. This process of consolidation entails gradual semanticization because the results of consolidation are semantically formatted memories stored in neocortical structures. The navigational theory proposes that we view memory systems as navigational systems. Just as spatial navigation proceeds from egocentric reference classes to allocentric reference classes over repeated exposures, episodic memories gradually shift to semantic memories over time and with repeated exposures. In sum, both theories predict that episodic memories gradually shift from being stored, retrieved, and formatted episodically to being stored, retrieved, and formatted semantically. Episodic memory forms a pipeline to semantic memory (see also Walker and Stickgold, 2010).

Semanticization puts pressure on a natural kind distinction between episodic and semantic memory by showing how episodic content continuously shifts to a semantic format. However, there is pressure in the opposite direction as well– pressure from the semantic to episodic direction. What I will call "the composition argument" has it that episodic memories must be composed by binding together abstracted semantic memories. In particular, Renoult et al. (2019) propose that an episodic memory is composed of, "a conjunction of familiar concepts and episode-specific information (such as sensory and spatial context)" (pg. 1046) (see Neisser, 1981). Renoult et al. (2019), by drawing on trace transformation theory (Sekeres et al., 2018), argue that semanticization of episodic memories result in gist and schema representations. These

representations, when activated during (episodic) recall, enlist various semanticized, depersonalized memories as well as perceptual and spatial representations that are bound together.[50] There can be more or less perceptual detail and semantic facts depending on current task demands (see also Irish, 2019; Vatansever et al., 2021). In short, the "what", "when", and "where" components stored in an episodic binding could only be derived from semanticized abstract facts.

As an example, when you remember your birthday party last year, the view says that you activate semanticized representations of the birthday party (e.g., that your party was at a particular restaurant, that you got a bouquet of flowers, that various friends were there, that you had the shrimp tacos, etc.). These semantic representations further activate various episode-specific information (e.g., the taste and smell of the shrimp tacos, the way the restaurant looked, smelled, sounded, the way your friends looked, etc.). This admixture of representations gets bound together forming your coherent recollective experience (and if the contents are rich enough, the signal computed on the results of the recollective process generates the associated episodic phenomenology).

Together, semanticization and composition entail a tight, dynamic interplay between episodic and semantic memory. Specifically, episodic memories are semanticized via consolidation, but moreover, episodic retrieval involves activation and binding of various semanticized memories together with perceptual and spatial representations. The result is a cyclical challenge to a natural kind distinction that recommends replacing a dichotomy with a

[50] See also De Brigard (forthcoming) who argues that causalism and simulationism are not opposed to one another. Episodic memory is simulation based, but requires traces. De Brigard argues that traces, understood according to a variant of Hippocampal Index Theory (see Goode et al., 2020), are a set of conditional instructions or dispositions to reactivate cortical structures for simulation.

richer life cycle of mnemonic representations, featuring numerous transitional forms: throughout the semanticization of episodic memories, and throughout the composition of episodic memories from variously abstracted semantic, perceptual, and spatial representations (see also Irish and Vatansever, 2020).

There is an obvious rebuttal to these concerns: even if it's true that the natural kind boundary cannot be drawn with reference to content or phenomenology, it might nonetheless be the case that episodic and semantic memory are modulated by distinctive underlying psychological or neural mechanisms. If this were right, then perhaps apparent gradation in surface properties can be mitigated by appealing to different underlying mechanisms, in the same way that gold and iron pyrite ("fool's gold") can be distinguished by their underlying molecular compositions despite having many similar surface properties. After all, semanticization implies a mechanistic division– episodic memories "shift" to semantic formatting and semantically associated neural systems.[51]

2.4 Double Dissociation Arguments

To put some teeth on the rebuttal given above, it is worth considering perhaps the strongest argument for a natural kind distinction between episodic and semantic memory: double dissociation arguments. X and Y doubly dissociate if X can function typically without Y, and Y without X. For example, in Broca's aphasia, it appears that conceptual knowledge is intact, but syntactic knowledge is lost. On the other hand, in Wernicke's aphasia, it appears that conceptual knowledge is lost, but syntactic knowledge is intact. If this is right, then syntactic and conceptual

[51] Aronowitz (forthcoming), to her credit, offers some skepticism about a double dissociation between episodic and semantic memory, but this skepticism stops short of showing that no such mechanistic division can be drawn. Hence, a natural kind division in terms of mechanism is still up for grabs.

knowledge doubly dissociate. Some authors have argued that double dissociations entail natural kindhood. Indeed, double dissociations are taken to be the gold standard for drawing modular conclusions about cognitive ontology.

Tulving (2002) suggested that we ought to look for double dissociations of episodic and semantic memory to decide on their kind status. Indeed, many lesion studies have been proffered as demonstrating such dissociations. Hippocampal lesions seem to produce loss of episodic function while preserving semantic function, whereas anterior temporal lobe lesions seem to produce loss of semantic function with intact episodic function. Hence, episodic and semantic memory are categorically distinct kinds.

However, there are reasons to be skeptical that such arguments are sound. In particular, Van Orden et al. (2001) have argued that double dissociations don't tell us much about modularity unless modularity is presupposed. Double dissociations require "pure cases"– that is, in the case of episodic and semantic memory, completely intact episodic abilities, with complete loss of semantic abilities (and vice versa). What Van Orden et al. (2001) showed is that there is no theory-neutral way to determine if some particular lesion is a pure case, and so cannot be used to arbitrate taxonomic disputes about kinds of memory (see Renoult & Rugg, 2020). In the case of the episodic-semantic distinction, there have long been doubts that particular lesion patients count as pure cases of dissociation, as supposedly pure episodic lesion patients also show subtle semantic deficits (e.g. McKoon, Ratcliff, and Dell, 1986), and patients with semantic dementia also show episodic-related deficits (e.g. Irish et al., 2012).

I am trying to show that semantic memory is very likely constructive, just like episodic memory, because episodic and semantic memory share underlying mechanisms. Hence, the double dissociation argument that would be sufficient to block my argument is more narrow in

scope. In particular, the defender of double dissociation just needs to show that the retrieval mechanisms of episodic and semantic memory are distinct. This is because construction is (primarily) a property of retrieval. So, is there evidence for distinct *retrieval* mechanisms? Do episodic and semantic retrieval mechanisms doubly dissociate?

Again, lesion studies have been used to argue that episodic and semantic memory have distinct retrieval mechanisms. Patient HM and other amnesiac patients have been used to show that the hippocampus mediates episodic retrieval, but not semantic retrieval. I will have more to say about hippocampal contributions to semantic retrieval in section 3. For now, I will say that it is far from clear whether amnesiac patients have spared semantic retrieval. On the other hand, semantic dementia and Alzheimer's patients have been used to show that loss of semantic retrieval spares episodic retrieval. Irish et al. (2012) cast doubt on this view. The authors tested these patients' episodic recall and episodic future thinking abilities. They found that patients with semantic dementia have relatively preserved episodic recall for *recent* past events, but deficits in episodic future thinking. Alzheimer's patients had paired deficits for episodic recall and episodic future thinking. This finding (among others, see La Corte et al., 2021; Paulin et al. 2020), suggests that the retrieval mechanisms for episodic and semantic memory are not purely dissociable. Indeed, it suggests continuity between these memory systems. In the following subsection, I will build upon the foundation laid here by mounting additional evidence for transitional gradation of episodic and semantic memory systems.

2.5 Transitional Gradation of Underlying Mechanisms

Semanticization and composition arguments show transitional gradation at the representational level. That is, episodic representations grade into semantic representations with a proliferation of transitional forms in between. I will here review evidence that the mechanisms

responsible for realizing these representations also grade into each other. That is, I will show that episodic and semantic memory systems (as opposed to representations), also exhibit transitional gradation. This presents a more robust and impervious challenge to a natural kind distinction because surface-level (representational) transitional gradation can be accommodated by appealing to a distinction in the underlying mechanisms. However, if the mechanisms themselves grade into one another, then no such categorical distinction can be made. Any attempt to do so will ultimately be arbitrary, meaning that the distinction can only serve heuristic purposes (Buckner, 2016; Gentry and Buckner, forthcoming).

Episodic memory, as reviewed above, has been characterized as mental time travel. This metaphor suggests that we might mentally traverse time in the same way we mentally traverse space. Indeed, the hypothesis I favor says that episodic and semantic memory are subserved by a domain-general system that supports representing and navigating relations among various kinds of stimuli, including space, time, events, semantic relations, and even dominance relations. For example, in cognitive mapping, agents frequently need to re-route to their goal when familiar routes are obstructed. The ability to flexibly re-route is subserved by spatial monotonic orderings of locations and landmarks in an integrated map-like representation. In episodic memory, the ability to mentally travel back in time to previously experienced events is subserved by the same system–one that monotonically orders events by temporal indices.More specifically, episodic memory and mental time travel abilities are exaptations of medial temporal lobe structures (especially the hippocampus and entorhinal cortex) possessing a more general function of arranging stimuli along ordinal monotonic dimensions and allowing agents to flexibly navigate around those representational spaces. This view fits broadly into more general trends in cognitive

ontology arguing that brain circuits are routinely reused for many different domain-specific functions (Anderson, 2010).

To further complicate the idea that different memory systems are dedicated to distinct mnemonic functions, Renoult et al. (2019) motivate the "composition argument" (reviewed above) by citing evidence for marked overlap in the neural correlates of episodic and semantic memory (see also Tanguay et al., 2023). Overlapping regions include: parahippocampal region, middle temporal gyrus, ventral parietal region, and midline frontal and posterior regions. Given the hippocampus' role in relational binding (Duff et al., 2020), it is somewhat surprising it is missing from the list. However, Solomon et al. (2019) found that the hippocampus redeploys relational binding mechanisms in semantic cognition. In particular, they found that the hippocampus represents semantic distances (e.g., cosine similarities in distributional models) between words in multidimensional semantic space.

It is generally accepted that the hippocampus codes for spatial relations using theta oscillations, but why would the hippocampus code for *semantic* relations if it were a dedicated episodic system? There is a rich history of viewing conceptual knowledge as organized according to similarity relations in multidimensional space (Gardenfors, 2004; Jones and Smith, 1993; Smith and Heise, 1992). Here, convex regions of regions of space correspond to concepts, and the distances between convex regions represent semantic similarity. The further the distance, the more dissimilar. This view has found renewed vigor in word embedding models where cosine similarities between words in multidimensional space represent semantic similarity (Boleda, 2020). Semantic similarity, according to this story, can be represented as a spatial relation (more on this later).

Further, Gazes et al. (2023) provide evidence of a domain-general magnitude representation system that responds to (at least) physical magnitudes, space, time, and dominance relations. Importantly, because of their shared underlying ordinal features, these magnitude representations can be extended to learned orders (acquired through, e.g., transitive inference). Gazes et al.'s account predicts that there should be common processing across various domains.

Focusing on spatial representation, the most popular story in mammals involves place and grid cell interactions in the medial temporal lobes to construct map-like representations of the environment (Moser et al., 2008). Here, place cells respond to locations and bind together various snapshots taken from different egocentric frames, grid cells represent the locations of these bundles with respect to one another in spatial dimensions by linking them to a spatially organized array. Visible spatial landmarks play an important role in establishing these links. The grids are anchored by landmarks, and the same landmarks are visible from different egocentric viewpoints. Just as Gazes et al. (2023) predict, the place and grid cell system seems to be redeployed to represent and reason about time (Cassasanto and Boroditsky, 2008; Eichenbaum, 2017; Rolls and Mills, 2019; Rueckemann et al., 2021). In particular, entorhinal "ramping cells" seem to have firing rates that track temporal order of events (Aghajan et al., 2023; Bright et al., 2020).[52] Moreover, some psychologists have argued that "temporal landmarks" play a similar role to spatial landmarks in anchoring a system of temporal representation in children (Shum, 1998; Tartas, 2001). A variety of other functions for grid coding have also recently been studied,

[52] Aghajan et al. (2023) suggest that ramping cells are grid-like cells in that their activity constitutes a fourth dimension (time) in topological representations of the environment.

such as “social place cells” that encode for abstract position in a social hierarchy (Omer et al., 2018).

In summation, the argument is that if episodic and semantic memory continuously grade into one another at the mechanistic level, then a natural kind distinction between episodic and semantic memory (likely) cannot be drawn. I have provided evidence that episodic memory and mental time travel abilities are exaptations of medial temporal lobe structures. In particular, I think recent neuroscientific evidence supports the view that mental time travel is subserved by a domain-general relational ordering system that deploys place and grid cells to construct map-like representations that are arranged along ordinal and monotonic dimensions. Landmarks (be they spatial or temporal) play an integral role in navigating these representational spaces because they help link together various points in these spaces. In the next section, I will be motivating the consensus view that episodic memory is constructive.

3. Episodic Recall is a Constructive Enterprise

Recall that my more general argument is that if episodic and semantic memory share underlying mechanisms and that episodic memory is constructive, then we have good reasons to think that semantic memory is constructive too. I have defended the first conjunct of the antecedent in the previous section. I turn now to defending the second conjunct: that episodic memory is constructive. To be sure, the composition argument by Renoult et al. (2019) (reviewed above) is a constructivist view of episodic memory, but that argument is relatively recent. There is, however, a rich history of viewing episodic memory as constructive. It will be worth looking at this history and the evidence supporting the view.

3.1 Constructive Episodic Recall

The received view of episodic recall is that it is constructive (Aronowitz, 2019; de Brigard, 2014; Michaelian, 2011, 2016, 2022). By this, I mean that episodic memory is not archival–episodic memories are *not* stored as informationally complete files and retrieved as such. Rather, a "trace" of the originally experienced event is encoded and stored. At retrieval, the trace is activated which triggers activation of perceptual, motoric, affective, and spatial representations to "fill in" the details in a probabilistic way. Today, most of the debate is over the nature of memory traces– a debate I will return to below.

The main reason a constructive view of memory is the received view is that a multitude of empirical phenomena seem incompatible with an archival view. In particular, memory distortions are pervasive in humans. If memory is archival, so the argument goes, we should not observe such widespread distortions. Michaelian (2011), drawing on Alba and Hasher (1983), list four kinds of memory distortion that seem incompatible with an archival view:

> 1. *Selection*: Only certain incoming stimuli are selected for encoding.
>
> 2. *Abstraction*: The meaning of a message is abstracted from the syntactic and lexical features of the message.
>
> 3. *Interpretation*: Relevant prior knowledge is invoked.
>
> 4. *Integration*: A holistic representation is formed from the products of the selection, abstraction, and interpretation process (pg. 325).

An example of potentially all four of the above properties, dates back to the early twentieth century psychologist F.C. Bartlett (1932).[53] Bartlett studied students' recollections of a

[53] For a detailed review see Roediger and McDermott (2000).

story they had been told, “The War of the Ghosts”. The students would systematically misremember the details of the story. Bartlett observed that the students seem to be reinterpreting the story using background knowledge (interpretation). The students would store a “kernel” of the story (selection, abstraction), but it would not make sense on its own. So they would embed this kernel in a context that made sense to them, but it did not faithfully correspond to the story (integration).

Another example of a memory distortion is the *relatedness effect*, “if people experience a series of items that are strongly related, they will tend to remember other (nonpresented) items as having occurred if these nonpresented items are strongly related to those that did occur” (Roediger and McDermott, 2000, pg.151). In 1959, psychologist J. Deese discovered that after being exposed to a list of 12 words, subjects will often report being exposed to nonpresented words that are strongly related to the list. For example, subjects will wrongly report being exposed to “sleep” when exposed to words like, “bed”, “awake”, and “rest”. Underwood (1965) calls this phenomenon *implicit associative responses* (or IARs). Underwood thought that when subjects were exposed to a word, they also unconsciously represented semantically related words. Michaelian (2011) thinks the explanation is that gist representations are stored which are sufficiently abstract so as to be compatible with nonpresented, yet semantically related words.

Another kind of memory distortion is confabulation. Garry and colleagues (1996) wondered: when people imagine events, are they more likely to judge that the imagined events actually happened to them? To test this, they asked subjects to vividly imagine an implausible childhood event that did not happen to them. Two weeks later, the subjects were given a sheet with a list of childhood events. Some of the events actually happened to the subjects and some did not. One of the events listed was the event that the subject had imagined two weeks prior.

The subjects were asked to rate the probability that the events happened to them in their childhood. Garry et al. found that subjects were more likely to judge that the imagined event had happened to them compared to controls.[54] The researchers call this phenomenon *imagination inflation*.

Of course, there is also semanticization and composition (reviewed above). Semanticization involves abstraction because episodic memories migrate to semantic memories. Composition involves interpretation and integration because the rememberer binds together various semanticized memories and integrates them with perceptual, spatial, and affective representations.

The empirical support for the pervasiveness of memory distortions poses difficult questions for the archival view of episodic memory. The archival view is committed to faithful encoding and preservation of memories. However, memory distortions seem incompatible with such a view. So episodic memory is constructive. If this is right, and most episodic memory researchers think it is, then what exactly is stored? And what does the constructive process look like? In answering these questions, it will be helpful to consider episodic recall and recollection from a computational perspective.

3.2 What is Stored? What is Constructed?

Felipe de Brigard (2022) has recently described the constructive process as a specific kind of computational problem:

> [T]he computational problem our memory system is trying to solve is a variant of what is known as 'an inverse problem': the challenge of determining, given a particular effect, what its cause must have been. Given the noisy and incomplete nature of the information the reconstructive process starts off with, the result is going to heavily rely on

[54] Similar results were found in Heaps and Nash (1999) and Hyman and Pentland (1996).

> completions that are highly probabilistic and dependent on background experience and conditions of recall (pg.7).

The "noisy and incomplete information" the process starts with is often called a memory "trace".[55] What is a memory trace though? Let's consider two views: the gist view and the pointer view.

Both the gist view and the pointer view agree that a memory trace is a content-bearing representation (cf. Vosgerau 2010). They disagree about what the content of the trace is. In short, the gist view says that a trace is a highly abstracted, summary representation of the target event. The pointer view, on the other hand, says that a trace is a set of conditional instructions to reactivate cortex. Moreover, both these views agree that cues activate traces which result in cascades of cortical activation constituting the recollective experience of the target event. I won't be taking a stand on either of these views because it's not entirely necessary for the more general argument I am trying to make. Again, I am interested in showing how if episodic and semantic memory mechanistically grade into one another, and episodic memory is constructive, then we have good reasons to think that semantic memory is constructive too. If both the gist view and the pointer view agree that memory is constructed at retrieval, then that's all I need for my argument.

---

[55] Some constructivist views deny that anything is stored (Michaleian 2016, 2022; Vosgerau 2010). For the sake of space, I am setting aside these kinds of views and focusing on constructivist views that posit traces. However, I think there are really good reasons to doubt no trace views. For one, if episodic recall is best understood as a kind of inverse problem, then it's not clear how it can be solved without a trace. But secondly, if nothing is stored, then it seems like there is no difference between imagination and memory. Imagination can be seen as a process of simulating experiences that one has not had before. But certainly memory is not like this. By definition, to remember something is to recall an experience one had. A trace is needed in order to preserve the distinction between imagination and memory. Indeed, Michaelian explicitly rejects the distinction in endorsing his no trace constructivism. This, I think, is a mistake (see also de Brigard 2022).

However, I will point to de Brigard's (2022) skepticism about the gist view because it will be relevant to how we should interpret the semantic pointer architecture that I will be discussing as a proof of concept for the view on offer here. De Brigard notes that if traces are abstract, compressed, summary representations, then a decompression process is necessary to explain the activation of cortex that constitutes the recollective experience. However, it is totally unclear whether decompression algorithms are biologically plausible. In particular, traces are alleged to be stored in the hippocampus, but there is no evidence to date that the hippocampus decompresses representations. De Brigard instead opts for the pointer view. As he describes it:

> [W]hen one experiences a certain event during encoding, the experienced content is instantiated in a particular representational vehicle, in the form of a hippocampal-neocortical network in the brain. Consolidation increases the probability of the nodes in the network to coactivate given the right cue. When such a cue is presented in the retrieval context, the coactivation among units of the network starts to propagate toward the hippocampal [trace], which does not contain explicit contents but rather the conditional instructions to reactivate the rest of the pattern of activity (2022, pg. 16).

This concludes my motivation for constructive episodic memory. For my part, I have motivated both of the premises in my overall argument. I have shown that episodic and semantic memory mechanistically grade into one another and I have shown that episodic memory is constructive. This entails, so I contend, that we have good reasons to think that semantic memory is constructive too. In the next section, I will review evidence that I think confirms this conclusion. In particular, I will argue for the view that, just like episodic memory, semantic memory is not archival. Semantic memory does not store a list of propositionally encoded statements that are ready for retrieval. Instead, semantic memory stores traces that are instructions to construct concepts.

4. Semantic Memory is Constructive

At this point, this reader might be skeptical that semantic memory is constructive. Indeed, they might agree with my premises, but nonetheless want to reject the conclusion. They might do this by reasoning as follows: well sure, semantic memory grades into a constructive episodic memory system, but that's compatible with semantic memory being archival. The reader might demand that I show positive evidence of constructive mechanisms from episodic memory being shared by semantic memory. In this section, I will attempt to do just that. Here's how I will proceed: since De Brigard (2022) focuses on episodic memory traces being stored and reactivated in the hippocampus, I will review some evidence that semantic memory implicates the hippocampus. I will then review some evidence that the hippocampus makes on-line contributions to semantic processing. Finally, I will close with some more general considerations in favor of a constructive semantic memory system.

4.1 Hippocampal Contributions to Semantic Processing

As discussed in section 2.4, Solomon et al. (2019) found that the hippocampus codes for semantic distances between words (semantic similarity; see Boleda 2020). Duff et al. (2020) in commenting on this study, say that:

> These data are striking as they suggest a role for the hippocampus in tracking and representing the relations among words in semantic memory in a manner that is similar to how the hippocampus tracks and represents relations in physical space and events in episodic memory (pp.11-12).

Recall that Michaleian (2011) lists four kinds of memory distortions that indicate constructive processes: selection, abstraction, interpretation, and integration. Episodic memory involves all four of these kinds of distortion. In particular, at recall, episodic memories show

signs of interpretation and integration. As the trace is activated, it employs cortex to "fill" in the spatial, perceptual, motoric, and affective details of the target event which results in an integrated recollective experience. Importantly, with episodic memory, interpretation processes can modify the recollective experience on-line. As De Brigard (2022) notes:

> [W]hen a memory is reactivated, it becomes modifiable. Since every act of retrieval is itself an act of re-encoding, nodes that weren't part of the original pattern but that already have a higher baseline probability of being coactive, are now more likely to getting [sic] included in the original pattern of activation after reconsolidation (pg. 16).

Semantic processing can be modified on-line too. In particular, the hippocampus makes on-line contributions to semantic cognition. Piai et al. (2016) presented subjects with incomplete sentences and they had to guess the missing word. There were two conditions: constrained and unconstrained. In the constrained condition, the sentence stem semantically constrained the possible missing word (e.g., "She locked the door with the ___."). In the unconstrained condition, the sentence stem did not semantically constrain the possible missing word (e.g., "She walked in here with the ___."). There was a delay after the sentence presentation, and then a picture of the missing word was presented to be named by the subjects. The crucial part was that the researchers measured hippocampal theta oscillations (which are known to code for spatial relations, and which Solomon et al. found were used to code for semantic distances). Piai et al. (2016) found that:

> In the constrained condition, all patients demonstrated increased theta power at this keyword compared to the preceding word, a pattern that was not present in the unconstrained condition. These results demonstrated that the hippocampus contributes to tracking and building semantic associations across words (Duff et al. 2020, pg. 12).

Here, we have evidence that the hippocampus is involved in representing relational information between words in multi-dimensional semantic space (see also Covington and Duff, 2016; Hilverman and Duff, 2019; Irish et al., 2012; Rubin et al., 2014). Further, these data show that the hippocampus makes on-line contributions to semantic processing. Connecting this to De Brigard's pointer view of memory traces: it could be that semantic memory traces are pointers stored in the hippocampus, and that the hippocampus represents the semantic similarity between the traces. Further, these relational representations can make on-line contributions to semantic processing, much in the same way that the hippocampus does for episodic memories.

The idea that the hippocampus houses episodic and semantic memory traces is not surprising given the argument from transitional gradation of underlying mechanisms. It is even less surprising when we consider semanticization processes. Semanticization involves a shift from episodic coding and formatting to semantic coding and formatting. If the mechanisms responsible for episodic and semantic memory grade into one another, then it makes sense that episodic and semantic memory traces would be stored together. However, there is more evidence that these traces are stored together. McKoon and Ratcliff (1986) demonstrated that episodic information is automatically and almost instantaneously (150-200 milliseconds) activated by semantic information in lexical decision tasks. In other words, semantic representations automatically and rapidly prime episodic representations. Strikingly, this was true even when there was an extremely low probability that the episodic information would be relevant. For example, in one of their experiments, there was only a 1/12 chance that an episodic prime-target pair would be tested. One plausible explanation for the automatic and rapid priming between episodic and semantic information is that they are stored together, allowing for rapid and automatic (co-)activation. Indeed, McKoon and Ratcliffe are actually responding to Tulving's

(1983) claim that episodic and semantic memory can be distinguished by processing speeds. Tulving claimed that episodic contents are accessed slowly and strategically, whereas semantic contents are accessed rapidly and automatically. McKoon and Ratcliffe went the extra mile by showing that not only can episodic contents be accessed rapidly and automatically, but that they can be primed by semantic representations.

4.2 Context and Construction of Semantic Representations

Finally, I'll close with some general considerations in favor of a constructive semantic memory system. So-called "concept contextualists" have argued for a constructivist view of semantic memory (Barsalou, 1987; Binder and Desai, 2011; Cassasanto and Lupyan, 2015; Davis and Yee, 2020; Dove, 2022; Löhr, 2017; Mazzone and Lalumera, 2010; Michel, 2020; Reilly et al., 2016; Yee and Thomspon-Schill, 2016). These theorists think that we should not view concepts as invariant structures stored in long-term memory because contextual demands are too variable for invariant concepts to accommodate. Here, I will briefly review a sampling of some of the issues that have motivated concept contextualism.

First on the list is emotional recognition. Gentry (forthcoming) has defended a constructivist view of emotion concepts based on Moebius syndrome patients' performance on face-based emotional recognition tasks. Moebius patients suffer from bilateral facial palsy which precludes them from making facial expressions. One popular view of emotional recognition is the facial feedback hypothesis which says that a necessary condition on successful emotional recognition is simulating facial expressions. This hypothesis predicts that Moebius patients will be deficient on emotional recognition tasks.

However, Gentry reviews studies with Moebius syndrome patients showing conflicting results: in some studies the patients show deficits, in others they do not. Gentry then diagnoses

these conflicting results as an issue with experimental design. In the studies where the patients were deficient, emotion labels were not used, but when the patients were not deficient, emotion labels were used. Gentry argues that a constructivist view of emotion concepts predicts this pattern of results. This argument has two main parts: firstly, Gentry argues that emotional experience is more rich than making facial expressions– it involves lots of other kinds of interoceptive and exteroceptive sensations. Thus, emotional concepts are richer than the facial feedback hypothesis predicts. Indeed, simulating facial expressions might be one way to deploy an emotion concept, but that does not exhaust the ways to deploy that same concept. Hence, Moebius patients should be able to simulate emotional experiences even if they cannot simulate facial expressions. Secondly, psychological constructionism about emotions says that language acts a cue to construct emotion concepts (Barrett, 2017; Lindquist and Gendron, 2013). The idea is that emotion labels serve as direct routes to emotional knowledge which allow Moebius patients to construct simulations of emotional experiences.

Next on the list is polysemy. Consider the following sentence: “Will drank the can and crushed it on his head”. Here, there are two occurrences of “can”. The second occurrence is masked due to anaphoric binding (the anaphor “it” is successfully bound to the first occurrence of “can”). However, these occurrences pick out different objects: the first occurrence picks out the liquid in Will’s stomach while the second occurrence refers to the crushed hunk of metal on his head.

Polysemy, understood as an instance of lexical ambiguity in which “words have a single meaning that can be modulated to fix distinct denotations depending on context”, pushes us to think that concepts might have single meanings with various distinct denotations (Quilty-Dunn, 2021; pg.164). Indeed, with the productivity and generativity of natural language, words can be

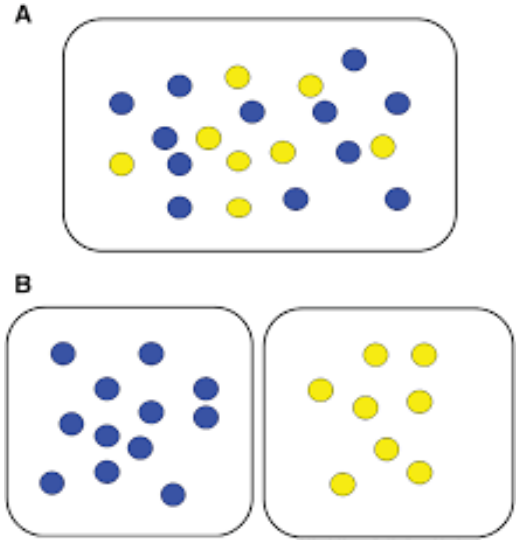
radically polysemous–single words can have multiple distinct denotations and/or every word could be made polysemous (see also Carston, 2021; Falkum, 2014; Liu, 2022; Pustejovsky, 1995; Recanati, 2017). As Hampton (2000) put it, "If concepts are to be tied fairly closely to substantive words…then concepts too must be amenable to many and varied contributions as components of thoughts" (pg.302) (cf. Brody and Feiman, 2023).

Paul Pietroski (2005) has defended concept constructivism arguing from polysemy. Consider another sentence (from Pietroski):

(F): France is hexagonal, and it is a republic.

The first occurrence of "France" picks out a geographical entity, but the second occurrence picks out a governing body. Pietroski suggests that "perhaps the meaning of an expression is an instruction for creating a concept from available mental resources" (pg.269). On this view, words get correlated with certain cognitive capacities, even one and the same word (polysemous words) get correlated with certain distinct capacities. For example, 'France' is correlated with thinking about certain spatio-temporal coordinates, and thinking about certain governing bodies. The two occurrences of 'France' in (F), and the grammatical features of (F), trigger those capacities. Put another way, lexicalization, for Pietroski, is a process in which diverse sets of mental representations get linked via language–linked to words. Because grammatical form contributes to meaning, and meanings are instructions to create concepts, grammatical form contributes to the recipe for concept creation. So the occurrences of 'France', together with the grammatical form of (F), jointly provide instructions to create concepts that resolve the polysemy.

Pietroski is not just motivated by philosophy of language. He also thinks there are empirical reasons to be a constructivist. Take a look at picture 8:



PICTURE 8. Example of stimuli from Knowlton et al. (2021).

In A, we see a mixture of blue and yellow dots whereas in B, we see spatially separated blue and yellow dots. Now consider the following two sentences:

(1). There are more blue dots.

(2). Most of the dots are blue.

There are exactly 12 blue dots and 8 yellow dots in each of A and B. So sentences (1) and (2) should pattern the same truth values, and plausibly there should not be a preferred matching between the sentences and the pictures. That is, each sentence describes each picture equally well.

Knowlton et al. (2021) experimentally showed that these two claims come apart. While it is true that each sentence describes each picture equally, subjects prefer one picture-sentence match over the other. In particular, subjects prefer sentence (1) and picture B, and sentence (2) and picture A. Knowlton et al. suggest that sentence (1) invites a comparative reading which is

best represented by picture B, involving comparison between subsets. On the other hand, sentence (2) invites a proportional reading which is best represented in picture A involving a superset. If this is right, then the prediction is that subjects will match the sense of sentences (1) and (2) to the corresponding picture.

This prediction was experimentally confirmed in multiple experiments. When given one of the two sentences and asked to choose between pictures A and B, subjects made the predicted matches. This also happened in the reverse direction: given one of the pictures and asked to choose between the two sentences, subjects made the predicted matches. More interestingly, if the subjects were presented with one of the sentences and asked to draw the image they think best captures the sentence's meaning, subjects drew pictures in the same spatial configurations as A and B, and these drawings corresponded to the predicted matching. For example, if given sentence (1), subjects drew pictures that approximated picture B.

Knowlton et al. argue that what explains these results is that "more" and "most" can be used to describe the same stimulus (in truth-functionally equivalent ways), but that each are correlated with distinct mental operations. In particular, "representation of a subset–subset relationship (*more*) or of a subset–superset relationship (*most*)" (2021, pg. 140; original italics). If meanings are instructions to build concepts, then "more" and "most" provide specific constraints on the kind of representations to build.

The evidence I have reviewed here–that episodic and semantic traces are jointly stored in the hippocampus, that the hippocampus makes online contributions to semantic processing, and that context effects the construction of concepts– all point in favor of a singular, constructive episodic-semantic memory system. In addition to the argument given in the previous sections– that because episodic and semantic memory mechanistically grade into one another, and because

episodic memory is constructive, semantic memory is also constructive– I have provided additional evidence for my conclusion. In the next section, I will be discussing a recent cognitive architecture that is a proof of concept for the view espoused here.

5. Semantic Pointer Architecture as a Proof of Concept

I have argued that episodic and semantic memory are not categorically distinguished because the mechanisms of each continuously grade into one another. I then showed how we should think that episodic memory is constructive. In particular, that episodic memory stores traces which, when activated, reactivate cortex constituting the rich recollective experience. Finally, I argued that if both of these claims are true, then it follows that we have good reason to think semantic memory is constructive too. The previous section was an attempt to bolster that conclusion by looking at evidence that suggests that episodic and semantic memory traces are stored together in the hippocampus, that the hippocampus makes online contributions to semantic processing, and that there are more general reasons to endorse constructivism about semantic memory, having to do with the contextual effects on concept deployment. I now want to introduce a proof of concept for the view espoused in this paper– the semantic pointer architecture (SPA) (Blouw et al., 2016; Eliasmith, 2013; Thagard and Schroder, 2014).

Recall that De Brigard (2022) thinks the best interpretation of memory traces is the pointer view–that the content of a memory trace is a set of conditional instructions to reactivate cortex as faithfully as possible. The SPA thinks that all of our concepts in semantic memory are like this. In other words:

> To possess a concept is to be able to activate various sequences of neural states that correspond to things like visual and auditory simulations, expressions of natural language, and motor commands all centered on a single category (Blouw et al. 2016, pg.1134).

Semantic pointers, according to the SPA, is what allows one to activate neural states centered on a category. A semantic pointer is a representation containing instructions to activate these various neural states. For example:

> [T]he concept CAR can be understood as a semantic pointer consisting of spiking neurons that compress, point to, and expand into other populations of spiking neurons that contain a wide range of information in various modalities such as verbal and visual (Thagard and Schroder 2014, pg. 5).

One might need to call upon some encyclopedic knowledge of cars when giving a presentation. In this case, the CAR pointer would activate more propositional knowledge; however, in another context, one might need to recall the smell of a new car. In this case, the pointer would activate olfactory simulations (see Blouw et al. 2016, pg. 1135).

The SPA is equipped to accommodate the contextual variability that cognizers find themselves in. If symbol-like representations are needed, then linguistic representations will be activated. If, on the other hand, more perceptual representations are needed, then sensorimotor representations are activated. The whole range of representational contents and formats, it is alleged, can be accommodated by pointing to (reactivating) various neural populations. Moreover, this kind of representational flexibility is exactly what one would expect from a singular, constructive episodic-semantic memory system. Indeed, Irish (2020) reaches a similar conclusion:

> The mechanisms underwriting the representational contents of [memory] can thus be conceptualized as operating akin to a pendulum, the relative loadings of which can be differentiated to the extent that perceptual (episodic) or conceptual (semantic) elements are required…In the face of changing task demands, the pendulum may swing intermittently between episodic and semantic loadings (Irish 2020, pg. 459).

Semantic pointers are brought about through a process of abstraction similar, if not identical, to semanticization. Blouw et al. (2016) liken the process to the kind of transformations that take place in the visual hierarchy. The mammalian neocortex possesses a 6-layer processing hierarchy. As information from vision passes through the hierarchy, it is iteratively transformed and compressed with correlative reductions in neural population sizes. More specifically, the detection of contrast difference and orientation is registered in V1, borders and lines in V2, angles and colors in V4, all the way to entire objects in anterior inferotemporal lobe (see DiCarlo and Cox, 2007; Yamins and DiCarlo, 2016). Importantly, the output of the visual hierarchy (in object perception) is an invariant object representation (tolerant of variation and noise). Notice that as information ascends the visual hierarchy, representations become more abstract and compressed (Buckner, 2018; Dove, 2022).[56]

Now, it is not the case that the kind of compression SPA invokes requires biologically implausible decompression algorithms (as De Brigard charges the gist view of memory traces). This is because the SPA does not think that the content of a semantic pointer is a compressed summary representation of the category. Rather, the content is just a set of conditional instructions to reactivate cortex. In Blouw et al.'s (2016) words:

> Semantic pointers cannot meet all of the desired criteria [for being a concept] when considered in isolation. Recall that a semantic pointer is simply a vector encoded by the spiking activity in a population of neurons. This vector captures relations between a wide range of other representations, and it can be transformed in various ways to access these representations, but the vector itself does not possess anywhere near the full semantic

[56] The "abstract" has multiple senses (see Reilly et al. 2023). Here, I have in mind a kind of vertical sense of abstraction wherein objects are more abstract than edges and orientation, the basic level category that the object belongs to is more abstract than the object (e.g., DOG is more abstract than that particular dog), MAMMAL is more abstract than DOG, etc. (see also Dove, 2022; especially chapter 4). I also consider unimodal representations to be less abstract than heteromodal and transmodal representations because transcending unimodal representations involves abstracting away certain details. In other words, heteromodal and transmodal representations, while still modal, generalize away from the particulars of a singular modality– they are less modality specifically.

> content of an ordinary concept. It is better, then, to think of a semantic pointer as an entity that enables the occurrence of a concept rather than as an entity that is equivalent to a concept (pg. 1134).

The SPA is a proof of concept for a constructive semantic memory system because concepts are not stored, invariant structures. Rather, concepts are constructed from semantic memory traces (semantic pointers). Furthermore, semantic pointers themselves are derived from episodic encodings of perceptual, motor, affective experience, much like the process of semanticization. Lastly, while SPA proponents are silent on where semantic pointers are stored, we have independent reasons for thinking they are stored alongside episodic traces (as I argued in section 3.1).

6. Conclusion

In this paper, I have argued that semantic memory is very likely constructive. Indeed, I argued that episodic and semantic memory grade into one another at the mechanistic level. Hence, there are not two separate memory systems, but one singular episodic-semantic system. The constructive nature of this memory system lies in the fact that they both store memory traces, understood as conditional instructions to reactivate cortex in accordance with probabilistic constraints and contextual demands. I reviewed the semantic pointer architecture as a proof of concept for this view on offer.

Accepting such a view of memory and concepts leads to some potentially radical conclusions. It's worth pointing them out here as future questions to pursue. Firstly, if concepts are constructed, then it seems to follow that a thought about P at T1 could be different in content from a thought about P at T2. This is a problem because presumably my different thoughts about P are still about P. How can conceptual representations be repeatable, in the sense of having the

same content? This question is ultimately about concept individuation or the identity conditions for concepts. As far as I know, no concept constructivist has given a full treatment of this problem. Indeed, some are happy to admit that conceptual representations are constantly changing, so there aren't any identity conditions to give (Casasanto and Lupyan, 2015). Machery (2009) argues that we should dispense with the concept of "concept" altogether because it doesn't constitute a natural kind.

However, another option is to distinguish what we do with concepts from their identity conditions. For example, we could identify a concept with the stored trace (i.e., semantic pointer), but deny that what is constructed at retrieval is the concept. All a concept *is* is the stored trace, what we *do* with a concept, what representations we construct at retrieval, is a separate issue. This move is precisely what modern day conceptual atomists commit to (Brody and Feiman, 2023; Quilty-Dunn, 2021). However enticing this move seems, it faces its own set of problems. For one, an informational atom, or a semantic pointer, fails to have the required content to satisfy various desiderata for a theory of concepts. For example, categorization behavior seems to involve representing various properties of the target category. Indeed, this is precisely what Blouw et al. (2016) had in mind when they said that a semantic pointer on its own cannot be a concept (pg. 1134).

Another potential issue for the constructivist is that sometimes (and perhaps all the time), we are interested in evaluating concept deployment. Did she correctly apply that concept? Is that judgment true? False? The classical computationalists (like Fodor, 1975) were very concerned with this question, and for good reason. Thoughts seem to be truth evaluable. However, on a constructivist view you might worry that truth evaluability is not possible, or at least highly problematic. For one, most constructivists are permissive with respect to representational format:

not all representations have a symbolic vehicle, some are imagistic, some are map-like, and some are embodied simulations (see Blouw et al., 2016; Dove 2009, 2022). How do we evaluate the truth of images or maps? These problems are not unique to a constructivist view, so I will set them aside. However, there is a potentially unique problem, and this is related to the first issue mentioned above. If there is no stability of meaning for one and the same concept, then one and the same concept application can be truth evaluated differently at different times. To be clear, this would not necessarily be a problem. One could apply DOG at T1 correctly and DOG at T2 correctly. The problem is that both applications could be intuitively correct applications, but because the content differs between the applications, we get different truth values. Again, I have not seen constructivists tackle this problem yet, although semantic internalism, contextualism, and holism remain live options (see Pietroski, 2005; 2018; Recanati, 2017; Werning, 2012).

However, another interesting option that hasn't been explored yet is to view conceptual representation as akin to scientific modeling. Wendy Parker (2020) has recently defended a view of scientific model evaluation that she calls the "adequacy for purpose" view. On her view, we should not ask whether a model is true. We should instead ask "can the model be used to do the job?". More precisely, she thinks we should view models as tools and evaluate the tools for adequacy with respect to a purpose as follows:

> A tool M is adequate-for-P if and only if using M in instance I results in the achievement of purpose P (pg. 461).

This view dispenses with the notion that a model needs to be representationally accurate, in the sense of truthfully modeling its object(s). In principle, a model can be highly idealized, but still achieve the purpose it was created for. Perhaps what constructivism asks us to do is to view

thought as modeling the world for particular purposes. We can do a better or worse job of this, but it's not clear that representational accuracy is a relevant dimension of evaluation when we are concerned with successful completion of goals or tasks (although sometimes it will be, e.g., when I want to accurately model the world). It's worth noting that this view fits nicely with how constructivists view episodic memory. The main motivation for constructivism about episodic memory is that memory is riddled with distortions and inaccuracies. Yet, we still manage to get around the world, pursuing our goals (largely) successfully.

I have not given an exhaustive list of the future questions and puzzles to be addressed by constructivists. Nor have I attempted to solve these issues, but hopefully this serves to show that there is much more work to be done.

REFERENCES

Aghajan, Z. M., Kreiman, G., & Fried, I. (2023). Minute-scale periodicity of neuronal firing in the human entorhinal cortex. *Cell Reports*, *42*(11).

Alba, J. W., & Hasher, L. (1983). Is memory schematic?. *Psychological Bulletin*, *93*(2), 203.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, *33*(4), 245-266.

Aronowitz, S. (forthcoming). Semanticization Challenges the Episodic–Semantic Distinction. *British Journal for the Philosophy of Science.*

Aronowitz, S. (2019). Memory is a modeling system. *Mind & Language*, *34*(4), 483-502.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. *Concepts and conceptual development: Ecological and intellectual factors in categorization*, *10139*.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, *15*(11), 527-536.

Blouw, P., Solodkin, E., Thagard, P., & Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive science*, *40*(5), 1128-1162.

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, *6*, 213-234.

Bright, I. M., Meister, M. L., Cruzado, N. A., Tiganj, Z., Buffalo, E. A., & Howard, M. W. (2020). A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proceedings of the National Academy of Sciences*, *117*(33), 20274-20283.

Brody, G., & Feiman, R. (2023). Polysemy does not exist, at least not in the relevant sense. *Mind & Language*.

Buckner, C. (2016). Transitional gradation in the mind: rethinking psychological kindhood. *The British Journal for the Philosophy of Science*.

Buckner, C. (2018). Empiricism without magic: Transformational abstraction in deep convolutional neural networks. *Synthese*, *195*(12), 5339-5372.

Carston, R. (2021). Polysemy: Pragmatics and sense conventions. *Mind & Language*, *36*(1), 108-133.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*,

106(2), 579-593.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In *The conceptual mind: New directions in the study of the concepts* (pp. 543-566). MIT press.

Cermak, L. S., Reale, L., & Baker, E. (1978). Alcoholic Korsakoff patients' retrieval from semantic memory. *Brain and Language*, *5*(2), 215-226.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(6699), 272-274.

Covington, N. V., & Duff, M. C. (2016). Expanding the language network: Direct contributions from the hippocampus. *Trends in cognitive sciences*, *20*(12), 869-870.

Davis, C. P., & Yee, E. (2021). Building semantic memory from embodied and distributional language experience. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(5), e1555.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, *191*, 155-185.

De Brigard, F. (forthcoming). Simulationism and Memory Traces. In Aronowitz, S. and Nadel, L. (Eds.) *Memory, Space and Time.* Oxford University Press.

De Brigard, F., Umanath, S., & Irish, M. (2022). Rethinking the distinction between episodic and semantic memory: Insights from the past, present, and future. *Memory & Cognition*, *50*(3), 459-463.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, *58*(1), 17.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, *11*(8), 333-341.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & cognition*, *24*, 523-533.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412-431.

Dove, G. (2022). *Abstract concepts and the embodied mind: Rethinking grounded cognition*. Oxford University Press.

Duff, M. C., Covington, N. V., Hilverman, C., & Cohen, N. J. (2020). Semantic memory and the hippocampus: Revisiting, reaffirming, and extending the reach of their critical relationship. *Frontiers in human neuroscience*, *13*, 471.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological review*, *111*(2), 524.

Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron*, *95*(5), 1007-1018.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. OUP USA.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Cambridge, MA: Harvard university press.

Gardenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*(2), 208-214.

Gazes, R. P., Templer, V. L., & Lazareva, O. F. (2023). Thinking about order: a review of common processing of magnitude and learned orders in animals. *Animal Cognition*, *26*(1), 299-317.

Gentry, H. (2023). Special attention to the self: a mechanistic model of patient RB's lost feeling of ownership. *Review of Philosophy and Psychology*, *14*(1), 57-85.

Gentry, H. (Forthcoming). Constructing Embodied Emotion with Language: Moebius Syndrome and Face-Based Emotion Recognition Revisited. *Australasian Journal of Philosophy.*

Gentry, H., & Buckner, C. (Forthcoming). Transitional gradation and the distinction between episodic and semantic memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*

Goode, T. D., Tanaka, K. Z., Sahay, A., & McHugh, T. J. (2020). An integrated index: engrams, place cells, and hippocampal memory. *Neuron*, *107*(5), 805-820.

Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic memory. *Neuropsychologia*, *38*(3), 313-324.

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological society*, *16*(5), 748-753.

Hampton, J. A. (2000). Concepts and prototypes. *Mind & Language, 15*(2–3), 299–307.

Heaps, C., & Nash, M. (1999). Individual differences in imagination inflation. *Psychonomic Bulletin & Review*, *6*(2), 313-318.

Hilverman, C., & Duff, M. C. (2021). Evidence of impaired naming in patients with hippocampal amnesia. *Hippocampus*, *31*(6), 612-626.

Hyman Jr, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of memory and language*, *35*(2), 101-117.

Irish, M. (2020). On the interaction between episodic and semantic representations–constructing a unified account of imagination. *The Cambridge handbook of the imagination*, 447-465.

Irish, M., & Vatansever, D. (2020). Rethinking the episodic-semantic distinction from a gradient perspective. *Current Opinion in Behavioral Sciences*, *32*, 43-49.

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. *Brain*, *135*(7), 2178-2191.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*(2), 113-139.

Klein, S. B. (2016). Autonoetic consciousness: Reconsidering the role of episodic memory in future-oriented self-projection. *Quarterly Journal of Experimental Psychology*, *69*(2), 381-401.

Knowlton, T., Hunter, T., Odic, D., Wellwood, A., Halberda, J., Pietroski, P., & Lidz, J. (2021). Linguistic meanings as cognitive instructions. *Annals of the New York Academy of Sciences*, *1500*(1), 134-144.

La Corte, V., Ferrieux, S., Abram, M., Bertrand, A., Dubois, B., Teichmann, M., & Piolino, P. (2021). The role of semantic memory in prospective memory and episodic future thinking: new insights from a case of semantic dementia. *Memory, 29*(8), 943-962.

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66-71.

Liu, M. (2022). Mental imagery and polysemy processing. *Journal of Consciousness Studies*.

Löhr, G. (2017). Abstract concepts, compositionality, and the contextualism-invariantism debate. *Philosophical Psychology*, *30*(6), 689-710.

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Mazzone, M., & Lalumera, E. (2010). Concepts: Stored or created?. *Minds and machines*, *20*, 47-68.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

McKoon, G., & Ratcliff, R. (1986). Automatic activation of episodic information in a semantic memory task. *Journal of Experimental Psychology*: Learning, Memory, and Cognition, *12*(1), 108.

McKoon, G., Ratcliff, R., & Dell, G. S. (1986). A critical evaluation of the semantic-episodic distinction. *Journal of Experimental psychology*. *Learning, Memory, and Cognition, 12*(2), 295-306.

Michaelian, K. (2011). Is memory a natural kind?. *Memory Studies*, *4*(2), 170-189.

Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. MIT Press.

Michaelian, K. (2022). Radicalizing simulationism: Remembering as imagining the (nonpersonal) past. *Philosophical Psychology*, 1-27.

Michel, C. (2020). Concept contextualism through the lens of Predictive Processing. *Philosophical Psychology*, *33*(4), 624-647.

Moser, E. I., Kropff, E., & Moser, M. B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, *31*, 69-89.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, *9*(1), 1-22.

Omer, D. B., Maimon, S. R., Las, L., & Ulanovsky, N. (2018). Social place-cells in the bat hippocampus. *Science*, *359*(6372), 218-224.

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, *87*(3), 457-477.

Paulin, T., Roquet, D., Kenett, Y. N., Savage, G., & Irish, M. (2020). The effect of semantic memory degeneration on creative thinking: A voxel-based morphometry analysis. *NeuroImage,* 220, 117073.

Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., & Knight, R. T. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences*, *113*(40), 11366-11371.

Pietroski, P. (2005). Meaning before truth. *Contextualism in philosophy: Knowledge, meaning, and truth*, *255*, 302.

Pietroski, P. M. (2018). *Conjoining meanings: Semantics without truth values*. Oxford University Press.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Quilty-Dunn, J. (2021). Polysemy and thought: Toward a generative theory of concepts. *Mind & Language*, *36*(1), 158-185.

Recanati, F. (2017). Contextualism and polysemy. *Dialectica*, *71*(3), 379–397.

Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: the dynamic multilevel reactivation framework. *Psychonomic bulletin & review*, *23*, 1002-1014.

Reilly, J., Diaz, M., Pylkkänen, L., Jefferies, E., Poeppel, D., Zubicaray, G., ... & Rodd, J. (2023). What we mean when we say semantic: A consensus statement on the nomenclature of semantic memory. *PsyArXiv preprint: https://osf. io/pre prints/psyarxiv/xrnb2*.

Renoult, L., Davidson, P. S., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends in cognitive sciences*, *16*(11), 550-558.

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: the semantic–episodic distinction. *Trends in cognitive sciences*, *23*(12), 1041-1057.

Renoult, L., & Rugg, M. D. (2020). An historical perspective on Endel Tulving's episodic-semantic distinction. *Neuropsychologia, 139*, 107366.

Robins, S. K. (2023). Kinding memory: Commentary on Muhammad Ali Khalidi's Cognitive ontology. *Mind & Language.*

Roediger, H. L., & McDermott, K. B. (2000). Distortions of memory. *The Oxford handbook of memory*, 149-162.

Rolls, E. T., & Mills, P. (2019). The generation of time in the hippocampal memory system. *Cell reports*, *28*(7), 1649-1658.

Rubin, D. C. (2022). A conceptual space for episodic and semantic memory. *Memory & cognition*, *50*(3), 464-477.

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological review*, *122*(1), 1.

Rubin, R. D., Watson, P. D., Duff, M. C., & Cohen, N. J. (2014). The role of the hippocampus in flexible cognition and social behavior. *Frontiers in human neuroscience*, *8*, 742.

Rueckemann, J. W., Sosa, M., Giocomo, L. M., & Buffalo, E. A. (2021). The grid code for ordered experience. *Nature Reviews Neuroscience*, *22*(10), 637-649.

Russell, J., Cheke, L. G., Clayton, N. S., & Meltzoff, A. N. (2011). What can What–When–Where (WWW) binding tasks tell us about young children's episodic foresight? Theory and two experiments. *Cognitive Development*, *26*(4), 356-370.

Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience letters*, *680*, 39-53.

Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological bulletin*, *124*(3), 423.

Solomon, E. A., Lega, B. C., Sperling, M. R., & Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences*, *116*(48), 24343-24352.

Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In *Advances in psychology* (Vol. 93, pp. 233-272). North-Holland.

Squire, L. R., Cohen, N. J., & Nadel, L. (1984). The medial temporal region and memory consolidation: A new hypothesis. *Memory consolidation: Psychobiology of cognition*, 185-210.

Tanguay, A. F., Palombo, D. J., Love, B., Glikstein, R., Davidson, P. S., & Renoult, L. (2023). The shared and unique neural correlates of personal semantic, general semantic, and episodic memory. *ELife*, *12*, e83645.

Tartas, V. (2001). The development of systems of conventional time: A study of the appropriation of temporal locations by four-to-ten-year old children. *European Journal of Psychology of Education*, *16*, 197-208.

Thagard, P., & Schröder, T. (2014). Emotions as semantic pointers: Constructive neural mechanisms. *The psychological construction of emotions. New York: Guilford*.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson, *Organization of memory.* Academic Press.

Tulving, E. (1983). Ecphoric processes in episodic memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *302*(1110), 361-371.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology, 53*(1), 1-25.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of experimental psychology*, *70*(1), 122.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove?. *Cognitive Science*, *25*(1), 111-172.

Vatansever, D., Smallwood, J., & Jefferies, E. (2021). Varying demands for cognitive control reveals shared neural processes supporting semantic and episodic memory retrieval. *Nature communications*, *12*(1), 2134.

Vosgerau, G. (2010). Memory and content. *Consciousness and Cognition*, *19*(3), 838-846.

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution. *Nature Reviews Neuroscience*, *11*(3), 218-218.

Werning, M. (2012) Non-Symbolic Compositional Representation and Its Neuronal Foundation: Towards An Emulative Semantics. In Hinzen, W., Machery, E., and Werning, M. (eds)., *The Oxford Handbook of Compositionality*.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, *114*(1), 152.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*, 616-641.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356-365.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic bulletin & review*, *23*, 1015-1027.