Non-Inferential Transitions: Imagery and Association

Jake Quilty-Dunn and Eric Mandelbaum

ABSTRACT

Abstract: Unconscious logical inference seems to rely on the syntactic structures of mental representations (Quilty-Dunn & Mandelbaum 2018). Other transitions, such as transitions using iconic representations and associative transitions, are harder to assimilate to syntax-based theories. Here we tackle these difficulties head on in the interest of a fuller taxonomy of mental transitions. Along the way we discuss how icons can be compositional without having constituent structure, and expand and defend the "symmetry condition" on Associationism (the idea that associative links and transitions are perfectly symmetric). In the end, we show how a BIT ("bare inferential transition") theory can cohabitate with these other non-inferential mental transitions.

§1: Kinds of Thinking

Thinking is not one kind of thing. The various forms of thinking studied in psychology include navigating with mental maps (Tolman, 1948; Camp, 2007; Rescorla, 2009), scanning and rotating iconic mental images (Shepard & Metzler, 1971; Kosslyn et al., 1978), sampling from a probability distribution (any Bayesian), detecting probable kin (Lieberman et al., 2007), making moral judgments (Mikhail, 2011), and doing simple arithmetic (Dehaene, 2011). Getting from thought A to thought B sometimes takes the form of a deductive, logical argument (Braine & O'Brien, 1998), but sometimes does not. One can think without utilizing full thoughts, for example. A thought like DONALD TRUMP IS THE PRESIDENT may lead you to think SOMEONE IS THE PRESIDENT; but, depending on your temperament, it might lead you simply to contemplate the abyss, which may just amount to thinking about the abyss as such and nothing more (that is, merely tokening the concept THE ABYSS).

A completed theory of thinking should exhibit full generality, capturing thought in all its various forms. One part of this theory concerns logical inferential transitions like the move from IT IS RAINING and IF IT IS RAINING THEN THE GAME IS CANCELLED to THE GAME IS CANCELLED. With respect to inference proper, we think there is something special about constituent structure.

Inferential transitions occur between thoughts based on rules that are built into the architecture of the mind and specify types of constituent structure (such as *If [P] and [IF P THEN Q] then [Q]*). We've argued elsewhere for this view of inference (Quilty-Dunn & Mandelbaum, 2017). Inference isn't everything, however, and may indeed account for a small part of our mental lives. In this article we intend to account for types of transitions that look to be inference-like but, according to our account, aren't genuine inferences. We'll also consider the structure of association in more detail, which sheds light not only on associative transitions but on their rule-governed foils, both inferential and non-inferential. Providing an adequate, coherent theory of thinking requires that we give an account of all these transitions. We aim to undertake this project, thereby providing a short (and surely incomplete) taxonomy of types of mental transitions here.

§2: Content-Specificity and Inference

Let's start with a quick overview of our theory. The fundamental case of inference, what we termed BITs (for Bare Inferential Transitions) was characterized as follows:

> The transition from state A to state B is inferential iff (i) A and B are discursive, (ii) some rule is built into the architecture such that A satisfies its antecedent in virtue of A's constituent structure and B satisfies its consequent in virtue of B's constituent structure (*modulo* logical constants), and (iii) there is no intervening factor responsible for the transition from A to B. (Quilty-Dunn & Mandelbaum, 2017, 8).[1]

Discursivity is central to this definition. This was, in part, to separate out two very different kinds of mental transition: associations and inferences. We'll say more about the difference between association and inference below. But one might wonder, aren't there mental representations that are not discursive? For instance, perhaps representations in perception (Carey, 2009; Burge, 2010; Block,

[1] We also distinguished BITs from RITs (*rich inferential transitions*), a fuller notion of inference in which the subject is also disposed to endorse the transition.

2014; Fodor, 2008) or mental imagery (Kosslyn, 1994) are iconic. Is it merely definitional that movements among them couldn't be inferential? We'll now look at how these kinds of transitions relate to inference.

Sentences and pictures have different sorts of structures. A sentence like "This is a blue square" involves discrete constituents that pick out an individual object ("This"), a color ("blue") and a shape ("square"). A picture of a blue square doesn't have this sort of structure—parts of the picture correspond to parts of the square (i.e., icons are *structure-preserving*), and the same part of the representation that picks out the object simultaneously represents its color and shape (i.e., icons are *holistic*). Icons thus lack the constituent structure distinctive of discursive representations, which tend to be neither structure-preserving nor holistic (Kosslyn, 1980; Fodor, 2007). We'll assume for the sake of argument that there are mental representations that are iconic in this sense, and that they're used online in perception as well as offline in mental imagery (Quilty-Dunn, forthcoming).

Icons figure in computations. They can be mentally rotated, such that the time needed to identify a match between two objects at different orientations increases as a function of increases in the difference in orientation (Shepard & Metzler, 1971). Icons can serve as inputs to categorization,[2] particularly via attentional selection of a subset of their contents (Sperling, 1960; Lamme, 2003). They can be scanned, such that sequentially retrieving information from spatially disparate parts of icons takes more time (due to more intermediating steps) as a function of the represented distance (Kosslyn et al., 1978). They also doubtless serve diverse functions only dimly glimpsed by cognitive science at present, such as their role in structuring long-term memory (Paivio, 1969).

Our present question is this: Do any of the computations that range over mental icons constitute inferences? And if not, what are they like? Certainly in the Helmholtzian sense in which

[2] We assume that categorization involves the application of a discursive concept to a perceived item (Mandelbaum, 2017); if that's wrong and the outputs of categorization can be icons as well (a possible interpretation of Prinz, 2002), then the computational role of icons is yet richer.

low-level perceptual computations constitute inferences, icons can be inputs to, outputs of, and brought to bear on inferences in perception. But if inference proper is limited to BITs, then icons may not be able to function in inferences. There is a very loose sense of 'inference' that ranges over any transition between states that solves an underdetermination problem (i.e., where the content of the output goes beyond the content of the input in some way). It is doubtful that much of interest can be said about the broad category that will also shed light on the particular type of transition involved in moving from the beliefs that p and if p then q to the belief that q. We don't deny that inference can occur unconsciously (on the contrary, it usually does) or that inferential transitions can range over "subpersonal" representations. We also don't deny that genuinely inferential transitions may occur in perceptual systems; we simply regard this as an empirical question, whereas the broad Helmholtzian notion of inference would make it trivially true that they do. Our approach is to take it that inference is a natural psychological kind of mental transition with paradigm cases (e.g., modus ponens) and explore whether instances of that kind occur in non-paradigm cases. One relevant difference between transitions in early vision and modus ponens inferences over beliefs is a difference in representational format. We turn now to transitions between icons that appear at first glance to be inference-like.

While some transitions involving icons don't seem at all like inferences—mental rotation, for example, might be useful in reasoning but seems qualitatively different than, say, *modus ponens*—others might. For example, some perceptual process might transition from an icon that encodes the color of an object and an icon that encodes the shape of an object to an icon that encodes the color and shape of the object. This kind of transition might be an instance of what Anne Treisman calls "feature integration" (Treisman & Gelade, 1980—though see Clark, 2004 for a non-iconic model). It

also resembles conjunction introduction, i.e., the transition from P and Q to P AND Q.[3] For present purposes, we're agnostic about whether this kind of process occurs (see Green & Quilty-Dunn, forthcoming for reasons to doubt the hypothesis that object-based feature integration is iconic). Supposing it does, however, should it count as an inference?

There is a quick argument which, presupposing the BIT view of inference, establishes a firm negative answer. BITs require rules to be built into the architecture that specify types of constituent structure. Icons lack constituent structure; therefore they fall outside of the scope of rules that specify types of constituent structure; therefore they cannot figure in BITs. Presupposing that something is an inference iff it is a BIT, icons therefore cannot figure in inferences.

Though this argument is, by our lights, sound, it fails to engage with the intuition that iconic transformations can implement paradigmatically inferential rules like conjunction introduction. If transitions between icons can in fact implement conjunction introduction, then any view that entails that iconic transformations are invariably non-inferential would seem at best to lack full generality and at worst to be arbitrary and false. We'll now consider some more principled reasons (i.e., ones that don't presuppose that something is an inference iff it's a BIT) to think that icons cannot implement conjunction introduction.

We understand the term "inference" to refer to a subset of the truth-preserving computations in the mind; that wider class excludes some processes, like associative transitions, but includes virtually all properly rule-governed ones. What marks out inference as we understand it from other sorts of rule-governed, truth-preserving computations is that inferences obey some *logic*. That is, the rules that govern them are logical rules, and they preserve truth in virtue of their form. That's not to say that the rules of human inference are identical to any logic taught in philosophy

[3] The case might better be modeled by introducing conjunction in the predicate terms, i.e., from X IS F and X IS G to X IS F AND G (though importantly not for Treisman, since feature integration *a la* Treisman & Gelade (1980) requires the introduction of a new object). We'll generally ignore this distinction.

courses. We suppose mental logic is idiosyncratic to humans (and possibly phylogenetically related ancestor species), partly for empirical reasons (Braine & O'Brien, 1998) and partly for *a priori* ones (Kripke, 1982).[4] What makes a rule logical is its formal character, i.e., that it abstracts away from all content except for the contents of logical operators like conjunction, disjunction, negation, and conditionals.

Their formal character allows logical rules to be relatively sparse in number and general in application. You don't need one inference rule to conclude from THE WEATHER IS BAD and IF THE WEATHER IS BAD THEN THE GAME IS CANCELLED that THE GAME IS CANCELLED and another to conclude from THE APPLE IS WAX and IF THE APPLE IS WAX THEN THE APPLE IS INEDIBLE that THE APPLE IS INEDIBLE. *Modus ponens* (and its psychofunctional equivalents—see note 4) is sufficient to generate the conclusion in both cases. The fact that rules of mental logic apply also to syntactically well-formed thoughts that are semantically inane such as "If there's a 3 then there's an 8" (Reverberi et al., 2012) and valid inferences that have unbelievable conclusions like "The feather is heavy" (Handley et al., 2011) provides reason to think that those rules are formal and therefore sparse and general.

What allows logical rules to be formal is that they specify types of structures that can be satisfied by representations. Suppose *If [F(X)] and [IF F(X) THEN G(X)] then [G(X)]* is an inferential rule. Since THE WEATHER IS BAD and THE APPLE IS WAX both instantiate the structure F(X), they can

[4] As we've argued (Quilty-Dunn & Mandelbaum, 2017), we don't see Kripkenstein as threatening to the psychological reality of rules as long as one grants a competence/performance distinction sufficiently robust to distinguish failure of conformity to a psychologically real rule (like *plus*) from conformity to a weird rule (like *quus*). And Kripke's insistence that the rule-following argument doesn't assume behavioristic limitations on mental architecture (1982, 14–15) seems to us to allow for such a distinction. But we accept the Kripkensteinian point that the functional properties of finite systems like human brains cannot distinguish rules with infinite scope, such as *modus ponens*, from weird rules that are, for such finite systems, extensionally equivalent, such as *modus schmonens* (which is equivalent to *modus ponens* except for conjunctions so large they cannot be thought by a human being in any possible world). Our account entails that all extensionally equivalent rules that describe 100% of transitions made in worlds where human beings think without performance errors are built into the architecture. Since these rules will thereby be truth-preservationally equivalent for propositions thinkable by humans, we don't see this entailment of our view as skeptical in any deeply worrying sense. Thus *modus ponens* and *modus schmonens* have equal psychological reality and human inference preserves truth no more or less than we pre-theoretically thought it did.

both function as premises in an inference governed by that rule. Satisfying the structural specifications of a rule is independent of the particular semantic values of (non-logical) constituents of the representation, so rules that only specify structural properties of representations are *ipso facto* formal rules.

Now we can go back to the case of an iconic transition that resembles conjunction introduction. We should note that since icons don't have constituent structures, they don't have constituents that function as logical operators (e.g., that express conjunction). But instead of harping on that point, we want to argue for the more interesting thesis that icons can't be governed by formal rules even independently of their inability to incorporate explicit logical operators.

The fact that icons lack constituent structures precludes them from satisfying structural specifications of logical rules. An iconic representation of a red square doesn't have separate constituents corresponding to redness and squareness, which means it doesn't share a constituent with an iconic representation of a red triangle. In that case, a red icon and a square icon cannot be combined into a complex red square icon; rather some rule has to map from the color and shape properties encoded in separate icons to a holistic icon that encodes both color and shape. A rule that says *If [F(x)] and [G(x)] then [F(x) & G(x)]* can't accomplish this task, since there is no constituent of the form F(x) in common between either of the input icons and the output icon.

This is not to deny that there can be rules that specify aspects of icons that allow for combination of contents. For example, a rule of the form *If [red*] and [square*] then [red square*]* can accomplish combination.[5] We also don't intend to deny that there can be rules that specify more abstract aspects of icons, such as *If [color*] and [shape*] then [color shape*]*, which ranges over all colors

[5] We use asterisks to denote aspects of icons.

and shapes.[6] But neither of these rules is identical to a formal rule such as conjunction introduction; they concern the combinations of features at different levels of abstraction.

The basic problem is that representations that don't have constituent structures will be governed by rules that are *content-specific*, in that they specify the identities of particular representations.[7] Consider rules that only range over syntactically atomic representations, e.g., *If A and B then C*. If this rule were not content-specific, then tokening any two representations would cause the tokening of any third representation with no limitations—i.e., thinking CAT and DOG would cause you to token any other atomic representation you can. For rules that govern atomic representations to be psychologically plausible, they have to specify particular representations. And in that case, they can't be formal rules and therefore can't govern inferences.

The case of icons is more complicated, however, since icons are not atomic. Icons break down into parts that correspond to parts of what they represent, and moreover each part encodes multiple features separately. A red icon with no shape information and a red square icon don't have a *constituent* in common, but they do have something in common—if we denied this fact, then there would be an explosion of primitives whereby every possible arrangement of features is a primitive icon that has nothing in common with any other icon.

What allows parts of icons to express multiple features at once is that they instantiate multiple properties at once, each of which corresponds to a property of what's represented. A part of a photograph might be red and square, and these properties of the part might correspond to the

[6] We're trying to be as non-committal as possible about what causes these features to be integrated; it might be that they're represented at the same location (in which case the rules should be enriched to reflect sameness of location) or simply that they're attended together.

[7] Note that content-specificity in this sense is compatible with being formal in a different, computational sense, i.e., that computation operates over symbols in virtue of their internal symbolic features and not what they actually represent out in the world (Fodor 1980). Here we mean that a rule that pertains only to the concept DOG is content-specific because it doesn't pertain to CAT despite DOG and CAT being structurally equivalent; it is compatible with this notion of content-specificity that the rule is "solipsistic", i.e., that it specifies the symbol DOG in a way that is completely blind to what DOG represents.

redness and squareness of what's represented. Another part might be red and triangular, corresponding to the redness and triangularity of what's represented. Since these parts are both red, they have a property (and content) in common. In this case the properties of parts are the properties of what's represented, which is surely not the case for mental icons. But what matters is that parts of mental icons have vehicular properties (i.e., properties of the representation itself, the "vehicle" of what content it carries) like red* and square* that enable them to express redness and squareness. How these vehicular properties should be individuated and how the compositionality of parts of icons should be modeled is an extraordinarily thorny issue we won't tackle here (for discussion see Quilty-Dunn, 2017). What's important for present purposes is that nothing about lacking constituent structure *per se* precludes a representation from combining features systematically, as long as it doesn't do so by means of composing discrete constituents (which the properties of redness, squareness, redness*, and squareness* are plainly not).

All this entails that icons can be compositional without having constituent structure, so there's no explosion of representational primitives. But the fact that icons compose features holistically (i.e., by means of properties of a single part rather than combining parts) means that their limited form of compositionality cannot be computationally exploited the same way the compositionality of discursive representations can. If a shape* value, an orientation* value and a location* value all hold of a given part of an icon, then accessing the shape* value requires accessing the orientation* and location* values as well. This means that a computational extraction of shape from an icon requires non-trivial work to differentiate the shape feature from its specific orientation, location, and any other holistically bound properties.

This *a priori* computational difficulty in extracting a single property from an icon has been noted by philosophers (Dretske, 1981; Haugeland, 1998; Fodor, 2003). It also makes sense of some of the most famous psychological effects involving icons. Take mental rotation. It's because the

shape and orientation of the object are represented by the same parts of the mental image that you can't immediately compare the shape of two objects at different orientations to see if they're identical. Instead, the orientation needs to match as well—that is, you have to mentally rotate the image of object B until its orientation* matches the orientation* of object A, and then see whether the simultaneously accessed shape* values concur as well.

The upshot of this discussion is that while icons can compose features without an explosion of representational primitives, the holistic manner in which they do so entails that computational processes have to access features in combination rather than individually. A consequence of this is that the rules governing those processes have to specify particular combinations without breaking them down into their primitive elements. Thus the rule that moves from a red* square* icon to a square* icon cannot simply be of the form *If AB then B*. A rule of that sort requires that B can be directly extracted from the complex AB, and the holisticity of icons simply precludes that sort of operation. An individual property can be extracted from an icon only by virtue of some (presumably highly complicated) intermediating process; the extraction cannot be a primitive computational operation. Not so for discursive representations: COW can be extracted directly from BROWN COW because it is a discrete constituent in its own right.

This moral applies in the other direction as well. There can be rules for composing discursive representations of the form *If [P] and [Q] then [P AND Q]* because P AND Q has P and Q as discrete constituents and ties them together via a conjunction operator. But this rule cannot hold over icons, since the computational process can't differentially specify the elements of the iconic counterpart of P AND Q (e.g., red* square*). Thus the feature integration process of taking a red* icon and a square* icon and delivering a red* square* icon cannot be a genuine instance of conjunction introduction but must instead implement some content-specific rule that maps red* and square* to red* square* under that specific description.

Note that while we've managed to avoid an explosion of primitive representational contents, we have run headlong into a mess of primitive computational rules. This would worry us if it weren't antecedently plausible. But as far as we can tell, perceptual psychology is up to its ears in massively complicated and proprietary algorithms—e.g., there's little reason to think that the computational rules implemented in deriving 3D shape on the basis of texture gradients are used for chromatic color constancy, or for anything else in the mind for that matter.

The content-specificity of rules for transforming icons also makes sense of an otherwise puzzling aspect of the mental imagery literature. Recall that the fact that orientation* and shape* have to be accessed together explains why you can't immediately compare the shapes of two objects represented at two different orientations. This means there needs to be some process that matches orientation* in order to match (or discriminate) shape*. But why does this process need to take longer the more intermediating orientation values there are—i.e., why does it have to be mental *rotation*? If the image of object A is at orientation-0* (i.e., perfectly upright) and the image of object B is at orientation-60* (i.e., at about two o'clock), why does the process that delivers an image of B at orientation-0* have to access the orientations* between 0* and 60*? The fact that shape* is bound up with orientation* doesn't by itself predict this result, since there could be a process that allows you to transform the orientation to whatever value you want. But this process doesn't underwrite the actual effect, since the effect is that reaction time increases as a linear function of the difference in orientation, and this strongly implies that the process represents the intermediating orientation values (otherwise what else would explain the linear increase?). If we suppose that rules for transforming icons are not content-specific, there's no principled explanation on offer for why the process can't immediately bridge any two orientation* values via some rule that abstracts away from specific values.

Now suppose that the rules for transforming icons *are* invariably content-specific. In that case, to go directly from orientation-60* to orientation-0* would require a specific rule for those specific orientations. Likewise for every combination of orientation* values, presenting an unwieldy computational burden. It would be relatively tractable, however, for there to be a content-specific rule for every specific orientation that simply delivers the adjacent orientation values. Thus the process could move from orientation-60* to orientation-59* via a content-specific rule and likewise for each pair between 60* and 0*. The content-specificity of rules governing iconic transformations thus explains the mental rotation results.

To sum up this section: inferences are governed by logical rules; logical rules are formal; formal rules specify structures independently of content; transitions between icons are governed by content-specific rules; so transitions between icons cannot be inferences.[8] We thus reject the claim that inferences occur in early vision not because we deny that early perceptual processes involve literal transformations over explicitly represented content in line with explicitly stored information but rather because the most plausible format for (many) such early representations precludes them from entering into inferential transitions. This applies to any transition featuring an icon as an input or output, even if the other representation being transitioned to (or from) is discursive.

So, for example, categorization on the basis of iconic inputs can't be a species of inference. There must be some content-specific rules that map types of icons to types of discursive representations. This point fits naturally with so-called "template-based" or "view-based" theories of categorization, on which categorization processes match stored unstructured viewpoint-

[8] We've discussed elsewhere some inference forms like semantic entailment (e.g., X IS RED to X IS COLORED) and probabilistic reasoning (Quilty-Dunn & Mandelbaum, 2018). The former seem to us to be genuinely inferential only when mediated by explicit knowledge of the connection (e.g., IF X IS RED THAN X IS COLORED) and therefore logical. The latter can be straightforwardly accommodated in our sort of framework (see, e.g., Goodman et al., 2015). A third form of inference is abduction, or "inference to the best explanation" (Harman, 1965). Abduction strikes us as a mess, one that has not been successfully captured by *any* extant theory and probably involves a maelstrom of different processes including task-relevant memory search, formulation of analogies, and probably some purely associative spreading activation and properly logical inferences. We put these forms of reasoning aside here.

dependent representations (which look to us to be icons) to inputs to see whether to apply a category (Ullman, 1996; Edelman & Intrator, 2001). On alternative "structural description" models (Biederman, 1987; Green, 2017) categorization processes instead convert iconic inputs into discursive representations that specify abstract structural features; if the incoming structural description accords with the stored description for some category, then the category is applied.[9] While the process of deploying a category on the basis of a structural description may be inferential—since both the category and the description are discursive, the transition may be governed by a formal rule—the part of the process that transforms an iconic input into a structural description must be content-specific. While on some theories icons play no role at any stage of perception (Pylyshyn, 2003), we assume that icons can at least occasionally function as inputs to categorization, which entails that categorization is not (or at least not always) a form of inference.

We're using 'icon' to refer to the pure icons that arguably figure in perception and mental imagery. There may be other sorts of icon-ish formats in the mind such as cognitive maps used for navigation. Rescorla (2009) argues that transitions involving maps only resemble logical inferences, and should instead be modeled differently due to their lack of explicit logical operators. We're happy to grant this distinction. However, we also think maps have constituent structures (Camp, 2007; Blumson, 2012), and that they are thus discursive representations that consist of an iconic representation of a spatial terrain organized discursively with other constituents of various possible formats (including markers that may themselves be discursive or iconic). It's thus possible on our account that iconic-discursive hybrids like maps could figure in transitions in virtue of rules that specify their constituent structures, and thus count as genuine inferences despite lacking logical operators.

[9] Biederman's theory is couched in terms of "geons," which are representations of types of cylindrical shapes of varying lengths, orientations, etc. Geons might look to be icons, especially when diagrammed (e.g., Prinz, 2002, p.140ff), but they are nothing more than descriptions, explicitly analogized to sentences by proponents of structural-description theories (e.g., Hummel, 2013, p.34).

§3: The Structure of Association

Transitions involving icons are interesting because they appear to be not quite inferential while also being clearly distinct from association. Association is a useful contrast with both inferential and many non-inferential processes because association is, in some deep way, *dumb*. Though we think a process such as mental rotation isn't strictly speaking an inferential one, for example, it nonetheless manifests complex computational intelligence.

We think it instructive to dwell on the nature of association, perhaps the most basic type of mental transition. Though the notion of association has a long and storied history (see, e.g., Mandelbaum, 2015; 2016), we think that previous analyses may have moved too quickly, and that the basics of the theory of association could use a reappraisal.

The term 'association' covers at least three different processes: associative learning, associative structures, and associative transitions. Associative learning refers to a paradigm in which one learns…something. What exactly is learned is part of the debate. One can put it neutrally by saying that one learns contigencies about the world, but a further question pertains to the structure of these acquired contingencies. Pure associationists (e.g., John Locke, David Hume, Ivan Pavlov, Anthony Dickinson, Celia Heyes) think what is acquired in an associative learning paradigm is an associative structure—a pair of mental representations that are structured associatively.[10] But what does "structured associatively" and its cognates ("instantiating an associative relation") amount to? Association is supposed to be the most basic relation that two mental representations can bear to one another. For example, Dickinson (1980, 85) describes association as "an excitatory link which

[10] N.b.: according to this way of cutting up the pie Skinner and Watson do not, strictly speaking, count as associationists because neither believes in mental representations; a fortiori they don't believe in associative links between mental representations and anything else. However, Skinner and Watson did think that associative learning causes the acquisition of an associative link between stimulus and response. In any case, given the demise of behaviorism we'll describe association from a mentalistic point of view.

has no other property than that of transmitting excitation from one event representation to another." Such a link is supposed to be maximally simple in that it is a brute connection between representations (one which is merely excitatory, not inhibitory).

The structures formed through association needn't be simple. It is reasonable to suppose that the semantic network of concepts in each person's mind is an enormous reticulated structure consisting of thousands of concepts linked together at varying associative strengths. What makes association dumb is thus not its simplicity, but rather its insensitivity to rational considerations. If you're simultaneously presented with a picture of Trump next to a picture of Rosa Luxemburg two hundred times, then you'll form an association between Trump and Luxemburg. Your knowledge of the fact that Trump and Luxemburg bear no sort of resemblance and have virtually no relation to each other won't prevent you from thinking LUXEMBURG when you think TRUMP. The only way to weaken the association would be an extinction paradigm, i.e., to see pictures of Trump without pictures of Luxemburg and vice versa.

One can therefore take an operationalized approach to the distinction between associative and non-associative transitions: the former can only be changed through extinction (and counter-conditioning) paradigms while the latter may be changeable in other ways, or not at all. This approach can be useful (Mandelbaum, 2015; Quilty-Dunn & Mandelbaum, 2017), but it isn't fully philosophically satisfying. Operationalism is an unacceptably superficial metaphysics of mind. *Pace* verificationists, operationalizations should be used as measures of underlying mental states and processes, not as specifications of their nature.

So, what is the nature of association? Associative transitions between mental states are mediated by a stored associative structure that links those mental states. A familiar point from centuries of rationalist critique of associationism is that an associative link does not have a propositional structure. Specifically, association is never sufficient for predication. Thinking SUGAR

and SWEET is not sufficient for thinking SUGAR IS SWEET. Thinking the full propositional thought requires some kind of semantically exploitable ordering, such that the meaning of SWEET is predicated of the meaning of SUGAR and not (necessarily) vice versa (Kant, 1781; Fodor, 2003). The predicative relation between constituents of propositional structures is thus asymmetric.

As noted, the argument that associative links don't suffice for propositional structure is quite well known. But the deeper point is that as classically understood, associative links don't merely have a non-propositional structure—they simply have no internal structure to speak of. An associative link is just the propensity of two representations to be activated together. It follows that, at least ideally, associative links are *symmetric*. That is, if A and B are associated, then (*ceteris paribus*!) activating either will cause the activation of its associate.

The symmetry condition on association leads to a problem, however, since most links between mental representations are not in fact symmetric. Thinking ULTERIOR leads to thinking MOTIVE quicker than thinking MOTIVE leads to thinking ULTERIOR. Some asymmetries can be explained away by differences in the other links between representations. Suppose that ULTERIOR is linked to very few concepts and MOTIVE is linked to many more, and suppose that the strength of associative spreading activation is a function not only of the strength of the link between two representations but also of the total number of linked concepts that have to be activated. In that case, the excitation off MOTIVE will be more diffuse than that stemming from ULTERIOR.

Another tool for explaining away associative asymmetry would be averting to inhibitory links. Such links would be outside the purview of association proper, however, and would thus be seen as extra performance constraints from an outside system affecting the true underlying nature of

association. Of course, differential linkage and inhibitory relations are by no means ad hoc posits—they are independently plausible parameters of theory construction for the mind.[11]

However, it seems that the asymmetry of association cannot fully be explained by appeal to linkage and inhibition. For one thing, there's no reason to think that MOTIVE actually *inhibits* ULTERIOR simply because it does not activate it to as high a level as ULTERIOR activates MOTIVE. Moreover, the fact that MOTIVE has more independent links doesn't seem like the only possible source of the asymmetry. It's plausible that the fact that the association is based in repeatedly hearing the phrase "ulterior motive" (where "ulterior" always precedes "motive" and rarely or never vice versa) is responsible for ULTERIOR activating MOTIVE faster (or to a higher activation level) than MOTIVE activates ULTERIOR. It seems plausible, furthermore, that this would hold even if they had the same number of links to other concepts (e.g., SALT and PEPPER may have roughly the same number of links to other concepts but there may be an asymmetry due to the fact that people hear "salt and pepper" much more often than "pepper and salt").

The plausibility of this sort of asymmetry creates a serious problem for the idea that association is a bare propensity for two mental states to be tokened together. Instead, there needs to be an ordering of the relation between the two representations, such that the link from SALT to PEPPER has a different strength than the link between PEPPER and SALT. In which case we see two options: 1) reject the empirical hypothesis that human beings actually harbor associative structures, or 2) augment the notion of association.

According to option 1, it turns out as a matter of fact that nothing, or nearly nothing, in the mind instantiates the bare relation posited by classical associationism. Though we are not associationists, we've happily granted that associations exist (Mandelbaum, 2016; Quilty-Dunn &

[11] For example, activating one of a pair of homonyms (e.g., [river] bank), will inhibit the activation of the other homonym (e.g., [savings] bank) Pylkkänen et al., (2006); Macgregor et al., (2015).

Mandelbaum, 2017). But other critics of associationism such as Jan De Houwer and colleagues (De Houwer, 2009; Mitchell et al., 2009) and C.R. Gallistel and collegues (Gallistel, 1990; Gallistel & King, 2009) have argued that even the associationist's favorite examples of learning (e.g., conditioning paradigms, learning in simple animals like insects, etc.) involve the acquisition of propositionally structured representations. De Houwer and colleagues hold that controlled reasoning is always needed to learn contingencies and that learning is never the consequence of mere automatic excitatory and inhibitory links. If asymmetry in structures like the one between ULTERIOR and MOTIVE are internal to the structure itself, then there may be reason to follow De Houwer down the path of rejecting associationism *in toto*.

However, denying the existence of associative links is probably an overreaction to asymmetry. For one thing, many associationists in the early stages of psychology explicitly thought of associative links as unidirectional (Thorndike [1932] referred to associative "polarity"; see also Rescorla, 1967). Indeed, early associationists puzzled over the existence of "backward association," i.e., the fact that presenting *a* before *b* led not only to a link from *a* to *b* but also from *b* to *a* (Cason, 1924; see also Ebbinghaus, 1885). Thus some associationists see asymmetry as a consequence of rather than a problem for associationism.

Furthermore, we suspect that the strong line taken by De Houwer and others is tied to learning; but learning is not the only form of evidence for the psychological reality of association. Lexical priming provides an independent source of evidence. Reading the word "doctor" will speed recognition of the word "nurse" (Meyer & Schvaneveldt, 1971) and reading "bug" will speed recognition of both "insect" and "microphone" independently of contextual disambiguation (i.e., reading "Ants are bugs" still activates MICROPHONE—Swinney, 1979). We have no idea how to propositionally model the link between DOCTOR and NURSE. Do you have to believe that doctors are nurses? Or that doctors often appear next to nurses? Aside from being blatantly *ad hoc*, these

hypotheses commit to empirical predictions about which representations mediate semantic priming that we strongly suspect wouldn't be borne out by the data. Instead, by far the most parsimonious explanation of (some forms of) semantic priming simply posits associative links between concepts and spreading activation that decreases in strength as it metastasizes (Anderson, 1983). These links can be understood as genuinely associative in that they are modulable through conditioning and that they are insensitive to rational evidence (as is clear in the BUG→MICROPHONE case).

Assuming association is worth saving, then if asymmetry is to be countenanced, perhaps the notion of association as the bare propensity of two representations to be co-activated ought to be augmented in some way. The most extreme reaction would be to simply throw out the idea of bare associative links and construe association as fundamentally structured. This reaction would alienate the spirit if not the letter of classical associationism—though some early associationists posited asymmetry, they never specified how the supposedly structure-free character of association allows for it. Not only have associationists described association as an unstructured pairing of representations, but this lack of structure plays a key role in the motivation for associationism. A core part of the appeal of associationism is its ontological simplicity, and its parsimonious ontology derives from the simplicity of the associative relation itself. Building structure into the basic associative relation thus drives a wedge between the theory of association and its aboriginal motivating idea.[12]

Moreover, supposing associations to be asymmetric creates a corresponding problem, viz., the problem of "backward association." It is generally true that even presentation orders that clearly

[12] It's perhaps puzzling why any associationist would ever have insisted on asymmetry given its tension with the idea that association is unstructured. But it's important to keep in mind the co-development of behaviorism and associationist models in the early twentieth century. Since many doubted the existence of representations, let alone associative links between representations, the one form of associative link that all associationists could agree on was the link from stimulus to response. And the notion of associative symmetry in stimulus-response associations makes little sense—being confronted with a ringing bell may associatively cause you to salivate, but salivation tends not to cause bells to ring. We thus suspect that behaviorist (and, more generally, operationalist) influence—not to mention the bare empirical fact that there are asymmetric transitions—caused some early associationists to regard association as asymmetric without seeing how uneasily that fits with their claim that association is unstructured.

favor a particular direction (A→B) will generate not only the desired "forward" link but also a "backward" link in the other direction, B→A (Asch & Ebenholtz, 1962; Hogan & Zentall, 1977). If association is asymmetric, then it's totally unclear why backward association would develop. Thus associationists who scrambled to appeal to independent factors to explain backward association (Cason, 1924; Storms, 1958).

Another reaction is to construe associative links as consisting of two layers. One is an unordered pairing of representations (i.e., the basic associative relation), and another is an ordered pairing with some strength. Thus if ULTERIOR and MOTIVE are truly associated, then activating either will guarantee the activation of the other; this is the bare associative link. However, while the bare associative link might guarantee whether or not a representation is activated (i.e., in a binary fashion), it might be silent on the *degree* of activation. The degree of activation might be modulated by a separate structure that specifies both direction and strength (e.g., ULTERIOR→MOTIVE, 0.8; MOTIVE→ULTERIOR, 0.2).

A significant theoretical cost of complicating associative structures in this way is that the structure of the transition from ULTERIOR to MOTIVE ends up being partially rule-governed, where the rule is something like *If ULTERIOR is activated, then activate MOTIVE to 0.8*. Nonetheless, this sort of transition could still be different from other rule-governed transitions in the following ways: (a) it's modulable through (and only through) conditioning; (b) it is blind to compositional structure and logical form (e.g., activating THAT IS NOT A BROWN COW will initiate all links that hold for both BROWN and COW); and (c) it is not truth preserving (e.g., if it's true that the table has salt on it, it may be false that it has pepper on it, but the associative transition proceeds anyway even if it has the kind of asymmetric structure on offer). This layered view faces a more significant problem in that it's not clear what explanatory work remains to be done by the bare symmetrical associative link.

We've been exploring these options in the hope of mapping out the logical space. But it is not in fact obvious that asymmetries like those discussed above are to be explained by appeal to association. In cases like the relation between ULTERIOR and MOTIVE, there is a complex concept ULTERIOR MOTIVE with a corresponding complex linguistic representation (viz., 'ulterior motive'), both of which have an ordering in which ULTERIOR/'ulterior' comes before MOTIVE/'motive'—and similarly for 'salt and pepper' and other cases. It may be that some non-associative structure or rule facilitates the activation of the complex conceptual/linguistic structure upon the activation of its primary constituent (viz., ULTERIOR/'ulterior') but not its secondary one (viz., MOTIVE/'motive'). In that case, ULTERIOR will cause the activation of MOTIVE both through the bare associative link between them and through activating ULTERIOR MOTIVE (and this transition may itself go through ULTERIOR activating 'ulterior', which activates 'ulterior motive', which activates ULTERIOR MOTIVE). But MOTIVE can only activate ULTERIOR through the bare associative link (which may have its own modulable strength as long as it's not unidirectional).

This difference would explain the asymmetry in activation strength and yet be wholly extrinsic to the association between ULTERIOR and MOTIVE, thus allowing us to avoid either discarding or complicating the associative relation itself. It does come at the cost of saying that presentation order facilitates the acquisition of non-associative structures (like ULTERIOR MOTIVE and 'ulterior motive') as well as associative links—but there's a wealth of evidence for that conclusion anyway (Mandelbaum, 2015; 2016), and it's particularly plausible in the case of complex linguistic constructions (the acquisition of which associationism is notoriously ill-equipped to explain). Linguistic structure is known to mediate the acquisition of associative links as well. For example, presenting two words together in a syntactically well-formed sentence forms a stronger automatic associative connection than presenting them in an ill-formed sentence; this effect can't be reduced to mere increased attention and cognitive resources but rather arises from the syntactic

structure itself (Prior & Bentin, 2008; and of course and instantiations of mere syntactic structures have their own associative effects).

Moreover, Asch and Ebenholtz (1962) rather plausibly argued that asymmetry arising from order of presentation was an effect of availability for recall—i.e., not because of any asymmetry in the associative link itself:

> When the task is that of paired-associate learning (or the learning of a set of *a-b* pairs), it is customary to require of *S* that he recall (i.e., anticipate) the *b* term (or the "response" member of the pair), whereas the *a* term (or the "stimulus") must be recognized and at best recited out loud. The test of backward association requires *S* for the first time to anticipate the *a* terms; since he has not had the opportunity to anticipate them previously, it would seem plausible that these items are less available to recall. Thus the study of backward association has not equated the conditions of availability, and often availability has systematically favored forward over backward recall.
>
> (Asch & Ebenholtz 1962, 139)

Availability for recall is a performance constraint that is highly context dependent. For example, if you spent yesterday thinking DOG over and over, then today you will tend to be better at recall tasks that involve dogs. This fact about availability for recall is independent of any associations DOG may have or acquire. Thus if you induce an association by repeatedly presenting the word 'dog' with 'New Jersey', then it will be easier to recall 'dog' upon reading 'New Jersey' than vice versa, but this fact is to be explained by differences in availability that are wholly extrinsic to the associative link itself (i.e., the fact that you happened to spend yesterday repeatedly activating DOG). Likewise, order of presentation may simply structure the availability of recall, causing a symmetrical associative link to manifest in asymmetric capacities for recall.

We can thus preserve the psychological reality of association as a bare symmetric relation as long as we posit multiple different interacting structures—some associative, some not—to explain the phenomena favored by associationists. Even semantic priming will turn out to be more than the activation of associative links. But viable forms of associationism are always replete with such complications. For example, while some forms of semantic priming seem purely associative (e.g., doctor–nurse), others are structured, such as priming that goes by way of superordinate–subordinate category structure (e.g., cousin–nephew) (Perea & Rosa, 2002; see McNamara, 2004 for a useful overview). As one would expect if one were positing distinct structures, these are dissociable: patients who display symptoms of Alzheimer's disease preserve purely associative lexical priming but lose priming that goes by way of category structure (Glosser et al., 1998).

The history of associationism is an object lesson in appending epicycles to an overly simple theory (as can be seen in, for example, stipulations made by Rescorla and Wagner [1972] in their famous model of associative learning). The continuing legitimacy of the notion of association as articulated in the earliest stages of psychology depends on these epicycles being mostly extrinsic to the associative relation itself; the associative relation should be complicated as minimally as possible and only as a last resort. We've noted a fundamental problem in asymmetric activation strengths, and outlined a number of possible explanations. The most likely outcome is that all these explanations are at least partially true and work in tandem: association is a bare propensity of coactivation but is also rarer than it's often thought to be, many classically associative phenomena are in fact undergirded by some non-associative as well as associative structures, and performance constraints like recall availability exert a significant impact on behavior.[13] Simplicity in the associative

[13] What kind of "bare propensity" is association? It's simply that there is no deeper psychological explanation for the propensity of the representations to be co-activated. Contiguity, resemblance and the like may serve to acquire the association, but once the association is acquired, the functional relation between the two is primitive from the standpoint of psychology. Activation spreading, for example, is not the sort of effect that is explicable in other psychological terms; instead, there are nodes in a network and levels of activation that spread associatively throughout the network. Why is it that activating one of a pair of nodes directly linked to each other will cause the other to activate? There is simply no

relation must be compensated for by complexity in other structures and conditions of acquisition. Human learning and memory are complicated, byzantine even, and that fact has to surface somewhere eventually.

§4: Conclusion

The mind is not simply a logical inference-making machine. We've outlined a number of non-inferential transitions, including iconic transformations, content-specific rules in central modules, and associations. This partial taxonomy leaves out a great deal of important mental processes (Bayesian processing being a glaring example, see Mandelbaum 2018). But it is intended to add some detail to a fuzzy sketch of how the mind might be structured and thereby provide some systematicity to our understanding of human psychology.[14]

References

Anderson, J.R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Asch, S.E., & Ebenholtz, S.M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, 106(2), 135–163.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

psychological-level story to tell. Whatever story there is to tell will be told in the vocabulary of sub-psychological reality (such as neuroscience, or whatever intermediate computational levels may connect neural activation to full-blown psychological functioning). None of this entails that the associative activation can't be modulated—associative links, by being primitive at the psychological level, are not therefore immutable rules built into the architecture. Counterconditioning and extinction can still suffice to modulate associative links. Thus associative links can have a characteristic functional role (viz., being so modulable) while nonetheless inhabiting a lower ontological level than inference and other properly psychological-level phenomena. Thanks to Anders Nes for the concerns that prompted these reflections.

[14] Quilty-Dunn's contribution to this project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 681422.

Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*, 89(3), 560-572.

Blumson, B. (2015). Mental maps. *Philosophy and Phenomenological Research* 85(2), 413–434.

Braine, M.D., & O'Brien, D.P. (Eds.). (1998). *Mental Logic*. Mahwaw, NJ: Psychology Press.

Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.

Camp, E. (2007). Thinking with maps. *Philosophical perspectives*, 21(1), 145-182.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Cason, H. (1924). The concept of backward association. *The American Journal of Psychology,* 35(2), 217-221.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*. NY: Oxford University Press.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, 37(1), 1-20.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. H.A. Ruger & C.E. Bussenius (tr.), 1913 (New York: Teacher's College, Columbia University).

Edelman, S., & Intrator, N. (2001). A productive, systematic framework for the representation of visual structure. In T.K. Leen, T.G. Dietterich, & V. Tresp (eds.), *Advances in Neural Information Processing Systems 13* (Cambridge, MA: MIT Press), 10–16.

Fodor, J.A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(1), 63–73.

Fodor, J.A. (2003). Hume Variations. Oxford: Oxford University Press.

Fodor, J.A. (2008). *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Gallistel, C.R. (1990). *The Organization of Learning*. Cambridge, MA: MIT press.

Gallistel, C.R., & King, A.P. (2009). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. West Sussex: Wiley-Blackwell.

Glosser, G., Friedman, R.B., Grugan, P.K., Lee, J.H., & Grossman, M. (1998). Lexical semantic and associative priming in Alzheimer's disease. *Neuropsychology*, 12(2), 218–224.

Green, E.J. (2017). On the perception of structure. *Noûs*. DOI: 10.1111/nous.12207.

Green, E.J., & Quilty-Dunn, J. (Forthcoming). What is an object file? *British Journal for the Philosophy of Science*.

Handley, S.J., Newstead, S.E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 28–43.

Haugeland, J. (1998). Representational genera. In his *Having Thought: Essays in the Metaphysics of Mind* (Cambridge, MA: Harvard University Press), 171–206.

Hogan, D.E., & Zentall, T.R. (1977). Backward associations in the pigeon. *The American Journal of Psychology,* 3–15.

Hummel, J.E. (2013). Object recognition. In D. Reisburg (ed.), *Oxford Handbook of Cognitive Psychology* (Oxford: Oxford University Press), 32–46.

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4(1), 47–60.

Kosslyn, S.M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S.M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

Kripke, S.A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, MA: Harvard University Press.

Lamme, V.A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18.

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, 445(7129), 727–731.

MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, *68*, 126-138.

Mandelbaum, E. (2015). Associationist theories of thought. The Stanford encyclopedia of philosophy.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3), 629–658.

Mandelbaum, E (2018). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language*. https://doi.org/10.1111/mila.12205

McNamara, T.P. (2004). *Semantic Priming: Perspectives from Memory and Word Recognition*. New York: Taylor & Francis Group.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227–234.

Mikhail, J. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge University Press.

Mitchell, C.J., De Houwer, J., & Lovibond, P.F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(2), 183–198.

Quilty-Dunn, J. (2017). *Syntax and Semantics of Perceptual Representation* (Doctoral dissertation, City University of New York).

Quilty-Dunn, J. (Forthcoming). Perceptual pluralism. *Noûs*.

Quilty-Dunn, J., & Mandelbaum, E. (2017). Inferential transitions. *Australasian Journal of Philosophy*, DOI: 10.1080/00048402.2017.1358754.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 241–263.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research,* 66(3), 180–194.

Prior, A., & Bentin, S. (2008). Word associations are formed incidentally during sentential semantic integration. *Acta Psychologica*, 127(1), 57–71.

Prinz, J.J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of cognitive neuroscience*, *18*(1), 97-109.

Pylyshyn, Z.W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT press.

Rescorla, M. (2009). Cognitive maps and the language of thought. The *British Journal for the Philosophy of Science,* 60(2), 377-407.

Rescorla, R.A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review*, 74(1), 71–80.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A.H. Black & W.F. Prokasy (eds.), *Classical conditioning II: Current Research and Theory* (New York: Appleton-Century-Crofts), 64–99.

Reverberi, C., Pischedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta Psychologica*, 139(1), 244–253.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *The British Journal for the Philosophy of Science*, *49*(4), 575–602.

Shanks, D.R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, 60, 291–309.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1–29.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645–659.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.

Storms, L.H. (1958). Apparent backward association: A situational effect. *Journal of Experimental Psychology*, 55(4), 390–395.

Thorndike, E.L. (1932). The Fundamentals of Learning. New York, NY: Teachers College Bureau of Publications.

Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.

Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychological* R*eview*, 55(4), 189–208.

Ullman, S. (1996). *High-Level Vision: Object Recognition and Visual Cognition*. Cambridge, MA: MIT Press.