Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-023-00719-4

Symbol and Substrate: A Methodological Approach to Computation in Cognitive Science

Avery Caulfield[1]

Accepted: 3 December 2023

Abstract
Cognitive scientists use computational models to represent the results of their experimental work and to guide further research. Neither of these claims is particularly controversial, but the philosophical and evidentiary statuses of these models are hotly debated. To clarify the issues, I return to Newell and Simon's 1972 exposition on the computational approach; they herald its ability to describe mental operations despite that the neuroscience of the time could not. Using work on visual imagery (cf. imagination) as a guide, I examine the extent to which this holds true today. Does contemporary neuroscience contain mechanisms capable of describing experimental results in imagery? I argue that it does not, first by exploring foundational achievements in imagery research then by showing that their neural basis cannot be specified. Newell and Simon's methodological position accordingly stands, even 50 years later. Computational — as opposed to physiological — descriptions must be retained to characterize and study mental phenomena, even as we learn high-level details of their implementation via brain data.

The cognitive revolution of the mid-20th century freed researchers of entrenched behaviorist conceptual constraints. Empowered to submit and to study mental operations, a number of programs in cognitive science have since developed considerably. By this, I mean that they have identified large numbers of well-formed, connected questions and made progress towards answering them experimentally.

As is well-known, cognitive scientists represent their achievements using computational models. What makes a model *computational* is not its realization on a laptop but its definition in functional, information-processing terms. One well-known research program that proceeded along these lines was described in Newell and Simon's 1972 *Human Problem Solving*, a foundational text in the field. They developed an account of the mental operations underlying problem solving behavior by carefully observing

✉ Avery Caulfield
acaulfi3@jh.edu

[1] Johns Hopkins University Department of Philosophy, Johns Hopkins University, Gilman Hall 3400 North Charles Street, Baltimore 21218, MD, USA

Published online: 26 January 2024

individual humans solving intricate problems in formal domains (e.g., chess playing). They introduce their attempt as follows (I apologize for the sexism):

> With a model of an information processing system, it becomes meaningful to try to represent in some detail a particular man at work on a particular task. Such a representation is no metaphor, but a precise symbolic model on the basis of which pertinent specific aspects of the man's problem solving behavior can be calculated. This model of symbol manipulation remains very much an approximation… This abstraction, though possibly severe, does provide a grip on symbolic behavior that was not available heretofore. It does, equally, steer away from physiological models… Perhaps the nonphysiological nature of the theory is not as disadvantageous as one might first believe, for the collection of mechanisms that are at present somewhat understood in neuropsychology is not at all adequate to the tasks dealt with in this book. We could not have proceeded to construct theories of human behavior in these tasks had we restricted ourselves to mechanisms that can today be provided with physiological bases (Newell and Simon 1972/2019, p.5).

Indeed, one finds very little reference to neurophysiology throughout this 1972 work, none whatsoever in the models of problem solving themselves. As they stress, their computational description was not a move away from psychology but an abstract description of "a particular man at work on a particular task" to which they were forced by an impoverished neuropsychology. Despite the clarity of the field's founders, much of the modern literature on cognitive science fails to appreciate this point. For instance, the "Computationalism" article in the *Oxford Handbook of the Philosophy of Cognitive Science* claims Newell and Simon were interested in "writing computer programs that simulate intelligent behavior without much concern for how brains work." Explanations at the computational and implementational levels were, according to Piccinini, considered "distinct and autonomous from one another" (Piccinini 2020, p. 182). But if we take the above quote seriously, no such commitment to autonomy exists. Had neurophysiology provided the tools necessary to describe high-level cognitive operations, Newell and Simon would have taken advantage of these tools.[1]

The merits of this information processing approach[2] – treating the mind as a quasi-insulated device whose functional properties can be explored in isolation from physiological models – can only be adequately assessed by examining in detail the successes and failures of research programs employing it. We will attempt a fragment of such an assessment in this paper, looking to Steve Kosslyn's work on visual mental

[1] This is not to deny that some authors avowed "in-principle" autonomy between the disciplines. They certainly did, though *Human Problem Solving* appears to have the correct idea.

[2] I will adopt Newell and Simon's loose usage of the term information processing for the purposes of this paper. I'm interested in exploring the "symbolic model[s] on the basis of which pertinent specific aspects of… behavior can be calculated." Whether these systems are truly *functionalist* or *information processing systems* in the technical senses of these terms is not germane to the questions I'm considering. See Piccinini's article (Piccinini and Scarantino 2010) for discussion.

imagery.[3] Kosslyn described his experimental program in detail in his 1980 *Image and Mind* and 1994 *Image and Brain*. The former openly proceeded at the "functional level" of description that I have characterized here (Kosslyn 1980, p. 124); at the time, Kosslyn noted his desire to eventually "develop the interface between the functional level and the neural substrate," but following Newell and Simon, the work proceeded "without regard to the underlying physiology"(Kosslyn 1980, p. 123-4). I'll show here that the latter 1994 work, despite appearances, does not deviate from *Image and Mind* in this regard.

Question: if we desired to describe in the most comprehensive manner possible our *current* understanding of high-level mental phenomena, what role would neurophysiology play in our account? In other words, how much has changed since Newell and Simon's 1972 statement of the relation between cognitive theory and the brain?

To answer this question, we might appeal to so-called *neuroimaging* techniques. These amount to sensitive instruments capable of measuring (or manipulating, e.g., TMS) physiological properties thought to be relevant to information processing (blood-flow, single-unit firing rates, etc.). The field which relies on these techniques as its primary means of accruing evidence is called *cognitive neuroscience. Image and Brain* made use of techniques from cognitive neuroscience[4] to develop the theory originally formulated in *Image and Mind.* In a 2001 paper titled "Neural foundations of imagery," Kosslyn goes so far as to claim that questions about visual imagery became "empirically tractable," able to be "tested objectively" with "the advent of cognitive neuroscience" (Kosslyn et al. 2001). If physiological measurements enabled true empirical scrutiny in imagery research, indeed, if the neural foundations of imagery were known, we might have moved beyond the abstracted, "mentalistic" theories described by Newell and Simon.

Others, however, take a very different perspective on the evidential role of neuroimaging. Consider Chomsky's response to a finding that event-related potential (ERP) readings "show distinctive responses to nondeviant and deviant expressions," even distinguishing four categories of linguistic deviance: "the current significance of the ERP studies lies primarily in their correlations with the much richer and better grounded C-R [computational-representational] theories." He goes on: within computational theory, "the five categories" with ERP correlates "have a place and, accordingly, a wide range of indirect empirical support; in isolation from C-R theories, the ERP observations are just curiosities, lacking a theoretical matrix" (Chomsky 2000, pp. 24-5).

Chomsky is, of course, not writing about visual imagery but generative syntax. These diverging accounts of neuroimaging's importance could be attributable to their studying different subject matters, and this is doubtless true to an extent: visual imagery *is* more amenable to evidence from cognitive neuroscience than generative syntax for reasons I will review later in this paper. However, I'll argue here that Kosslyn overstates the evidential role of neuroimaging even in his own field. Kosslyn's functional theory

[3] Imagery might first be likened to imagination, but seeing as the study of imagery has come to encapsulate involuntary activation of the visual system (as in, for instance, pattern recognition), we cannot rely on the colloquial term too heavily.

[4] In addition to improved behavioral measures and prevailing perceptual theory.

was developed on the basis of "objective" tests, and as Kosslyn himself pointed out, his 1980 computational model provides "nonmetaphorical explanatory answers to a wide range of questions, such as questions about why some tasks take longer than others, why some are more difficult than others, etc." (Smith and Kosslyn 1980). Furthermore, as will be discussed, key behaviorally-motivated features of the theory have endured severe empirical scrutiny over the course of the program; in some cases, the support is so substantial that these details have claim to permanent psychological knowledge.[5]

Despite the title of Kosslyn's 2001 article, no physiological model of imagery is presently known; the specified "neural foundations" amount to coarse localizations of processes whose characterization remains computational – that is, our knowledge of these processes can at present only be described in computational, as opposed to physiological, terms. Neuroimaging data have begun to shed light on large-scale features of how the brain implements Kosslyn's cognitive theory, but without the theory these data are of little evidential value—it's not clear what they tell us about. What's worse, Kosslyn's explicitly computational approach to studying imagery has largely been abandoned by contemporary researchers; Kosslyn was unable to locate a copy of any of his models when queried, even after contacting many collaborators. By neglecting the model, and thus the history of behavioral results motivating its features, modern imagery research loses considerable evidentiary force. Without good reason, it fails to engage with a history of interconnected experimentation that could significantly constrain answers to further questions about imagery.

If my positions on these matters are correct, then surprisingly little has changed since 1972 from an evidential perspective. The language of information processing remains the only medium by which detailed accounts of high-level mental phenomena can be expressed and developed. We are forced to describe mental phenomena computationally if we would like to describe them at all. Neuroimaging may support, develop, and localize components of cognitive theory, leading to real implementational understanding, but true "neural foundations" of imagery or other high-level mental phenomena are out of reach.

1 Kosslyn's Behavioral Work

Visual imagination is a commonly reported depictive phenomenological experience. It received systematic experimental treatment in Kosslyn's (1980) work *Image and Mind*, though this was neither the first nor the last inquiry into its properties (see, for instance, Paivio 1975) Kosslyn (1980). The highly successful inquiry into this subject has spanned more than 40 years, answering questions about the imagery system itself and how it enters into other operations in the mind. The two experiments I review here will illustrate each of these kinds of development. The first occurs relatively early in the program, asking whether visual images are *assembled from multiple distinct encodings* when formed, or generated *all at once* from a single underlying representation. Kosslyn conducted the second experiment – which sought to under-

[5] This is not to suggest that impermanent knowledge exists. A claim shown to be false was never known in the first place (Azzouni 2020).

stand how people conduct size-comparisons from memory – much later. Kosslyn had good reason to suspect that size-comparisons involved imagery and at least one other system, so the question became how these cognitive operations interact. Needless to say, none of these cognitive phenomena are observable to the eye. Only by extremely careful experimental design could Kosslyn gain access to the domain of unconscious, high-level mental operation. Let's see how he did it.

1.1 Early Generation Experiments

Consider two possible models of visual image generation: (M1) images are formed "all at once," as though pulling a picture from a scrapbook; (M2) images are assembled "piece by piece." There is considerable phenomenological evidence for M2: everyone who I have asked is capable of imaging a polar bear dancing in the desert, despite (presumably) never having seen anything like that before.[6] This suggests that people are capable of combining imagistic units creatively. But suppose we wanted to gain experimental access to this question; how would we do it? Kosslyn creates *conditionals* whose antecedents are our alternative hypotheses and whose consequents are distinct behavioral expectations. Were the consequents identical across the different hypotheses, the conditionals would be useless to us, for they would not discriminate alternative possible answers to our question.

Here's the conditional: "if the 'piecemeal retrieval' view is correct, then the simple amount of stored material will dictate how much time is required to form an image" (Kosslyn 1980, p. 99) . Under M2, then, increasing the amount of material that subjects need to image should increase generation time. If, on the other hand, M1 is the case, then the amount of stored material should not influence generation time. We would not, of course, expect complexity of an image to influence how long it takes to pull from a scrapbook.

Kosslyn designed a number of experiments that answer this question, the most solid of which I review here. Subjects learned to image drawings of animals, with the drawings divided into three groups. In one group, the animal was presented to subjects on a single page; in another, it was divided onto two pages (body on one, appendages on another); in a third, "the animal was divided into five sections, in roughly hierarchical fashion working from the body outward" (see Fig. 1) (Kosslyn 1980, p. 103) .[7]

Subjects learned to form a visual image of the animal when presented with its name. Those in groups 2 and 3 were never shown the entire animal, and so had to combine the units they had been shown into a single image. After being presented with a name, subjects formed an image of the animal and pressed a button when their image was fully formed (thereby measuring generation time). Kosslyn included a number of cross checks to ensure that subjects really were imaging the entire animal and not just a piece, but I have no time to review them here. For our purposes, what matters is that generation times "increased *linearly* with number of units" (Kosslyn 1980, p. 104). This is particularly striking in light of the fact that the same amount of visual material is present across groups, with the only differences being in how that

[6] "Image" here simply means "form a mental image."

[7] The experiment is described in greater detail in Kosslyn et al. (1983).

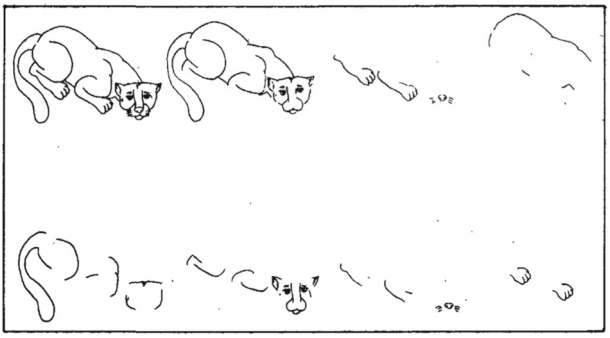

Fig. 1 Example stimuli from Kosslyn et al. (1983). The first group's stimulus is in the top left, the second's are the middle two on top, and the remainder are the third's

material is divided and presented to subjects.[8] These results are "difficult to explain" unless images are formed by amalgamating units from separate encodings, piece by piece (Kosslyn 1980, p. 104).

Notice what has just been achieved: we've gained access to an unobservable functional property of our cognition. We did so by designing an experiment whose results we expected to qualitatively differ under each of our alternative hypotheses. Our observations clearly discriminated these alternatives, leading us to the conclusion that mental images may be amalgamated from separate encodings in long-term memory. These results, among others, led Kosslyn to postulate mental operations like the "PUT" procedure, which "integrates a stored encoding of the appearance of a part" (long-term memory representation) "into a pattern already in the surface image" (Kosslyn 1980, p. 149). PUT, in a later version of the theory, is replaced by a shifting "attention window" which searches for a "foundation part"; "once it is found… a new image is formed" by activating a "pattern activation subsystem." Because this process is "iterative… repeated for each additional part or characteristic that is added to the image" the results described here are explained by the newer theory (Kosslyn 1994, p. 294, 389). The reader may complain that theory change suggests no progress is being made towards specification of psychological mechanism. The extent to which theory change is informed by experiment, as opposed to what is merely "in vogue" for non-empirical reasons must be examined on a case-by-case basis. To conduct such an examination would be outside the scope of this essay, but this complaint also misses something important about the experimental result just reviewed: we have *learned something permanent* about the imagery system, namely that "time to generate an image increases with increasing numbers of parts" (Kosslyn 1994, p. 294). Any theory of imagery which fails to account for this highly systematic finding demands elaboration; the results described here therefore place real constraints on any future explanation, constraints that cannot be ignored by any attempted assessment of progress in psychology.[9]

[8] This is what makes the present experiment so solid: it is impervious to possible effects of factors like image density.

[9] Importantly, the constraints imposed on explanation by any single experiment are far less severe than those of an entire program. The scope of Kosslyn's work is considerably broader than the isolated results I

1.2 Size-Comparison Experiments

The other experiment I'd like to review was conducted rather late in the *Image and Mind* program, focusing on the question of how human subjects conduct size-comparisons from memory; Kosslyn asks "How do you, the reader, decide which is larger, a mouse or a hamster?" then later the same question about a rat and an elephant (Kosslyn 1980, p. 349).[10]

Kosslyn notes that subjects commonly answer these two questions differently. Whereas the rodents provoke reports of visual imagination, subjects "just know" rats and elephants to be of different sizes (Kosslyn 1980, p. 349). The former suggests that subjects make use of quasi-pictorial mental representations to compare the sizes of the animals, as one would if actively perceiving those animals; the latter suggests the use of some kind of descriptive (propositional, categorical) model, like size tags—elephants may have a "large" tag, and rats a "small" tag. In addition to noting these reports, Kosslyn reviews a few experimental results suggesting that both imagistic and descriptive representations are used in these size-comparison tasks, including Paivio's demonstration that subjects could more quickly name the larger of two objects the greater their size disparity (called the "size-disparity effect") Paivio (1975) (Kosslyn 1980, p. 350-1). The size-disparity effect is taken to suggest the involvement of the imagery system because a parallel effect is observed in perception – people more quickly determine which of two objects is larger if their size difference is more pronounced.

The above constitutes suggestive evidence for the involvement of both depictive and descriptive systems in size-comparisons from memory. The following questions then arise: are both kinds of representations and their associated mental operations truly used? If so, in what way do they interact? A positive answer to the former question and any determinate answer to the second (by experiment) would both evidence the existence of quasi-pictorial mental representations and possibly the properties of the representational system itself. Let's examine some alternative possibilities for their interaction (or lack thereof) on these tasks.

One obvious possibility is that one of these two systems isn't used. We may conduct size-comparisons with *just mental images* (without propositional/categorical representations), or with *just propositions* (no images). Another possibility is that one system is used, and if/when that system fails, the other one is invoked. For instance, animals might be stored in the coarse categories SMALL, MEDIUM, and LARGE such that these size tags are immediately consulted when asked to perform size comparisons; if the animals are in the same category, then the imagistic system could be used to conduct more fine-grained comparisons where the categories fail. Call this model a "propositional-imagistic serial model" (P-I serial model). We might also consider an imagistic-propositional (I-P) serial model, whereby images are immediately formed, then used to generate size tags. Finally, we could use an imagistic-propositional parallel (I-P parallel) model, in which people simultaneously consult stored category and image representations, ultimately relying on whichever system generates an answer

review in detail here—the image generation literature alone has many robust findings (e.g., Kosslyn 1994, p. 294-5).

[10] The experiment is described in greater detail in Kosslyn et al. (1977).

more quickly. These are 5 distinct classes of models (see Kosslyn 1980, p. 351-4 for further details, though I will describe the expected behavior of each of these models in an experiment below).

Strikingly, Kosslyn designs an intricate experiment capable of discriminating between *all five* (!) of these models. Subjects learn to draw six simple stick figures, each a "different size and a different color" (Kosslyn 1980, p. 354). Subjects are first trained to draw a figure of the correct size when given a color, then learn to categorize the figures into two groups – small and large. Subjects are trained both to name the category a figure is in when given a color and to name the colors in a particular category. The experimental groups are distinguished by the extent of "overlearning" the category labels: one group is tested until they recall the correct category associations just two times in a row (200% group), and the other until they have done so perfectly five times in a row (500% group). "The subjects then are given pairs of color names, and are asked to judge as quickly as possible which name labels the larger figure. The results of primary interest concern pairs which contain stimuli that are (1) of similar or dissimilar sizes and (2) in the same or different categories" (355). Before I describe why each model predicts distinct results in this experiment, notice the following: when propositional models are used, differences in category overlearning between the 500% and 200% groups may play a role, but we don't expect this to affect imagistic models; when imagistic models are used, we expect to observe increasing size-disparity effects as the figures' sizes diverge, with no effects of category overlearning. The predictions of each model are thus as follows:

The **purely imagistic** model predicts that we will observe the size-disparity effect with "no effects of category membership," meaning overlearning the category labels should have no effect on the experimental results (Kosslyn 1980, p. 355).

One **purely propositional** model posits serial access to gross, then detailed information if it's required. This model can account for the size disparity effect because accessing and comparing the more detailed information required to make more fine-grained distinctions takes more time. While this model predicts size-disparity effects to arise when comparing between vs. within group size-discrimination times, these effects should not arise based on within-group variation in sizes, as the same level of detail must be accessed.[11] Because the overlearning procedure applies to the gross categories, this model predicts faster responses in the 500% than the 200% group in *both* the within and between category conditions, as the gross information needs to be accessed in both conditions. Another purely propositional model posits parallel access to gross and detailed information, predicting different-category decisions to be faster for the 500% group and same-category decisions to be the same in both groups (as overlearning the gross categories would not affect this same-category process).

[11] This took me a moment to wrap my head around. An example helped. If this model uses gross category information to distinguish the very largest and very smallest stick figures (so the largest in the large group, the smallest in the small group), this should take the *same amount of time* as distinguishing the second largest and second smallest, as in each case all that must be compared is gross category information. The same is true of comparisons within the detailed group; no size-disparity effects are expected for comparing the smallest and second smallest, as opposed to the smallest and third smallest figures. Needless to say, this is not so for the imagistic models.

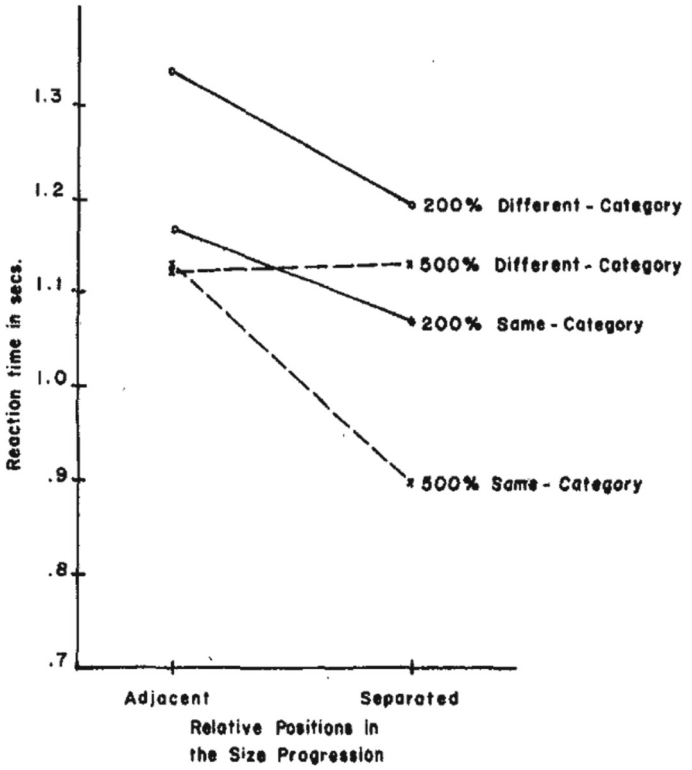

Fig. 2 Results from the size-comparison experiment Kosslyn et al. (1977), consistent only with the I-P parallel model

The **P-I serial** models make similar predictions to the purely propositional serial models, except that same-category judgments (when images are required, as propositional information no longer discriminates) should display size-disparity effects.[12]

In the **I-P serial** models, we expect no effects of category overlearning on the size-disparity effect, as imagistic (not propositional) representations are responsible for gross comparisons. See (Kosslyn 1980, p. 356) for details of the model.

Finally, the **I-P parallel** model. This model amounts to a "race" between the two submodels, with the faster being responsible for answer-generation. This model makes the peculiar prediction that, if differential category membership can be determined quickly enough (as in the 500% condition), then an answer can be generated propositionally "before imagistic comparisons are completed," thereby circumventing size-disparity effects when stimuli are in different categories (Kosslyn 1980, p. 356). If the tag retrieval process is slower, however, and the imagistic process outruns the propositional one, we expect to observe size-disparity effects for both same-category

[12] There is a latent assumption here that the relevant categories cognitively are the learned ones. The results we will observe render this assumption unproblematic.

and different-category stimuli. This is the only model that generates the highly specific prediction that size-disparity effects will be observable in all conditions except for the between-category 500% condition, and this is exactly the data that we observe (see Fig. 2).

1.3 Concluding Remarks on the Kosslyn Behavioral Experiments

The above experiments may impress the reader in their own right. In response to questions about the nature of image generation in the first case and question-answering from memory in the second, Kosslyn composed behavioral experiments whose clear results distinguished alternative accounts of underlying cognitive mechanisms. In the former case, we discovered that mental images are generated iteratively by assembling organized units, and in the latter case we learned that size-comparisons from memory are conducted by parallel imagistic-propositional processing. The conditionals underlying these experiments amount to instances of (fuzzy) MEASURES (Smith and Raghav's term) – a "specific equation licensing inference of values of a less accessible quantity from measured values of more accessible quantities" (Smith and Seth 2020, p. 424). I say "fuzzy" because these conditionals connecting behavior to cognitive mechanism are neither precise, nor theory-mediated, nor even quantitative. The conditional I reviewed above – "if the "piecemeal retrieval" view is correct, then the simple amount of stored material will dictate how much time is required to form an image" (Kosslyn 1980, p. 99) – amounts to a coarse-grained inference from underlying computational mechanism to overt behavior; it is validated by observation of multiple sources of clear data that evidence the same conclusion based on this conditional, which Kosslyn achieves and reviews alongside his presentation of both of the above experiments.[13]

Again, this alone may impress the reader, but understanding the strongest evidence in the Kosslyn program requires more scrutiny. The conclusions Kosslyn reaches themselves continue to be validated long beyond the completion of the experiment. This is so because Kosslyn repeatedly *took the results of previous work to be true* then *asked and answered* questions based on these assumptions. The ability to observe answers to presupposition-laden questions garners evidence for those presuppositions. *Image and Mind* is filled with "decision trees" that illustrate just this pattern of evidence accrual. Most obviously, both of the experiments just reviewed presuppose that there exist quasi-pictorial mental representations, asking questions about their generation and how they interact with other systems. Of course, Kosslyn gathered converging evidence for the quasi-pictoriality of images – via the scanning experiments (Kosslyn

[13] 1-5 in the previous section are similarly coarse inferences from underlying mechanism to overt behavior. More precision, closer to the kind familiar from developed sciences (Smith 2014), can be achieved by tuning cognitive theory to measured parameters within individuals, the approach of Kosslyn et al. (1984). This is surely more productive than the approaches to individual differences common in the psychology literature, namely developing survey-based measurements of "vividness," "control," and other adjectives whose cognitive correlate isn't clear.

1980, p. 36-52) – but the evidence achieved by continually presupposing and building on the results of these experiments is far stronger.[14]

Notice just how much must be presupposed for Kosslyn to specify the complex interaction of mental imagery with a descriptive system in the above size-comparison experiments! As the reader will expect by now, these presuppositions and the discovery that subjects used the I-P parallel model to conduct size-comparisons continued to be validated even after the conclusion of the above experiment, via their entrenchment in the computational model. Kosslyn went on to "suppose that propositional representations are processed at the same time as imagery representations, and that the relative speed of processing the two sorts of information determines which representation is in fact used," using this supposition to answer questions about how the imagery system is used to retrieve factual memories (Kosslyn 1980, p. 387).

Importantly, these presuppositions do not insulate further work from overturning them. Checks against our traveling down a "garden-path" pervade Kosslyn's work. The size-comparison experiments just reviewed constitutively presuppose the existence of the visual imagery system; our observation of the selective and systematic appearance of the size-disparity effect along the lines predicted by one of the models indicate that our assumption was correct (Smith 2014). As Kosslyn notes, "On cracked and crumbling foundations stands a shaky house" – enduring stability of results reassures us that our theoretical foundations are solid (Kosslyn 1980, p. 9).

The evidentiary value of constitutively presupposing prior conclusions in new research may lead one to expect widespread adoption of this strategy in psychology. This is unfortunately not the case. Looking to the psychological literature, one finds a large amount of clever experimental design in service of answers to individual, isolated questions. For instance, several linguists have pointed out that "research involving UG [universal grammar] is typically cast in terms of whether or not such an entity exists" Miller et al. (2016).[15,16] Researchers devote tremendous effort to statistical analysis of isolated results in an effort to quantify the evidence they achieve, doubtless because they are seen as the "gold standard" of evidence in psychology (Wagenmakers 2007). But rarely do psychologists conduct chains of experiments whose very design constitutively presupposes prior results, meaning the evidence achieved in Kosslyn's early work may be among the strongest in the history of cognitive science. Concerns about the validity of conclusions drawn from a particular experiment (in this case, Pylyshyn's (Pylyshyn 1984, ch. 4) can be answered by demonstrating the ways in which these conclusions continued to be tested indirectly, "*en passant*," in subsequent work (Smith 2010).

[14] Somewhat more subtly, to study image "generation" is to presume some distinction between stored representations and those being actively entertained.

[15] I don't make this point to berate developmental linguists in the generative tradition any more than they already have been (e.g., Christiansen and Chater 2008, 2009). Instead my point is to suggest that carrying on research in pursuit of UG's *properties*, as opposed to merely its existence, may be both evidentially productive and more effective at meeting criticism than the current strategy.

[16] Similarly, "Most research on predictive language processing in the last 15-20 years has focused on demonstrating that prediction is an important part of language processing. Much less research has been directed at establishing the mechanisms and mediating factors of anticipatory language processing" (Huettig 2015).

The observation that psychologists often fail to take advantage of this key evidentiary strategy is even more salient in the context of the oft-discussed *replication crisis* in psychological research: published "findings" often fail to replicate when researchers subject them to other tests (Shrout and Rodgers 2018). In a recent *Annual Review* article on the subject (Shrout and Rodgers 2018), recommendations for insulating future research against being overturned included improving statistical techniques, preregistering hypotheses, designs, and materials, reporting null results, and others. Absent was any mention of Kosslyn's research strategy, with its demonstrated history of unusual success. Smith and Kosslyn himself noted in a joint 1980 article the "common point in the philosophy of science that empirical research in a theoretical vacuum is likely to flounder." Unlike "a mature science, such as physics or molecular biology," cognitive psychology has "no established theory to rely on" as a means of guiding and constraining further research. They argue that Kosslyn's approach is "an appropriate response to this problem of doing research" absent established theory, an argument that has apparently been neglected by the modern literature.[17]

2 Kosslyn's 2001 Claims

The above results (and others, reviewed in Kosslyn 1980, Kosslyn 1994) are striking, particularly to those familiar with the experimental depth normally achieved in contemporary cognitive literature. Perhaps more striking is that they have largely been forgotten – or at least significantly underestimated – by modern researchers, even Kosslyn himself. In a paper published 20 years after *Image and Mind*, Kosslyn makes a number of assertions about the scope of the knowledge achieved with "psychological" (as opposed to neuroscientific) methods that I claim are highly misleading.

Kosslyn asserts that "philosophy and cognitive psychology" have "raised important questions about imagery, but have not made substantial progress in answering them. With the advent of cognitive neuroscience, these questions have become empirically tractable," revealing such phenomena as "the ways in which imagery draws on mechanisms used in other activities, such as perception and motor control." He goes on: "new neuroimaging technologies, especially positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), allow theories of imagery to be tested objectively in humans."

[17] Kosslyn's cumulative approach described here is not unique in the history of psychology. Consider, for instance, Hubel and Wiesel's early exploration of the critical period in cat and monkey visual development. They gained access to the properties of this period "by closing one eye at different ages and keeping it closed for several months or longer" then measuring, for example "the relative influence of the two eyes on single cortical cells" (Wiesel 1982). Later, via autoradiography measurements taken from a monkey whose monocular deprivation was switched from one eye to another (at 3 weeks), they discovered "the critical period is different for the two cell types. Whereas the critical period is over for the magnocellular input at 3 weeks, the parvocellular input apparently begins to lose its ability to expand at 6 weeks" (Wiesel 1982). Their ability to more finely resolve these details of the critical period (and others, e.g. "competitive mechanisms rather than disuse are prime factors in producing the changes observed under conditions of monocular deprivation.") is similarly enabled by large numbers of presuppositions (which are indirectly tested) about the visual system and its development.

In light of the results just reviewed, these are obvious understatements of the evidence achieved by pre-neuroimaging mental imagery research. We asked *and answered* tremendous numbers of questions about visual imagery before the development of neuroimaging technologies, going back at least as far as Mary Cheves West Perky's 1910 demonstration that perceptions can be mistaken for images, and certainly even further (Perky 1910). As Kosslyn notes at an earlier date, his 1980 model provides "nonmetaphorical explanatory answers to a wide range of questions, such as questions about why some tasks take longer than others, why some are more difficult than others, etc." (Smith and Kosslyn 1980).

I scrutinize Kosslyn's 2001 claims because I believe they express a widespread sentiment that behavioral work on computational problems (in "box-psychology") is somehow less rigorous (objective, informative…) than work that points a sensitive instrument at the brain. In fact, just the opposite is true from an evidential perspective for two reasons, one contemporary, one historical.

Most visibly, even advanced results in imagery (like the size-comparison experiment) rely on behavioral measures. These measures are legitimated insofar as they are capable of producing clear results that can be built on, as in the above case and many others.

Historically (and perhaps more importantly), behavioral results constitute the evidential basis of ongoing research. Above, we observed the cumulative nature of results in the study of visual imagery; because past results enter constitutively into subsequent successful research, early work *enables* further experimentation. The discovery that images are composed piece by piece, for instance, is among the most foundational results in imagery research. Thus, even if contemporary researchers had abandoned behavioral techniques in favor of measures from cognitive neuroscience (they have not), behavioral experimentation would still play an evidentiary role in the background so long as its conclusions continue to be presupposed in cognitive theory.[18]

In light of these observations, it's difficult to interpret which theories Kosslyn had in mind whose "objective" testing was enabled by fMRI and PET scanning. It's unlikely he was referring to his own computational theory of imagery (either the 1980 or 1994 version), as behavioral experiment figures centrally in its development and testing.[19] To address this question, we need to briefly consider the "imagery debates." In response to imagery research conducted by Kosslyn, Paivio, and others, critiques of the program surfaced. Zenon Pylyshyn's, the most well-known, attacked the invocation of visual images on the grounds of its failing to genuinely explain mental phenomena. For Pylyshyn, true explanation of the mental must be symbolic. While Kosslyn's theory was developed and implemented computationally (meaning it is symbolic in one sense), the units at its core are imagistic, not propositional.

Pylyshyn's limitations on explanation are purely stipulative. We have no evidence-based reasons to limit ourselves to propositional representations. Moreover, we do have considerable evidence that people employ depictive mental representations, discussed

[18] I will eventually point out that much contemporary research fails to engage with cognitive theory, to its detriment.

[19] It's a little misleading to refer to to theory development and testing as though they are independent. As I have stressed in this paper, theory components are subject to stringent test by subsequent theory development.

above. The scanning experiments provided initial converging evidence that people use depictive mental representations, and subsequent research overwhelmingly validated this conclusion *en passant*.

From an evidentiary perspective, then, Pylyshyn's worries are no concern, even if we limit our discussion to behavioral experiment. Interestingly, Kosslyn does not refer to the behavioral in his responses to Pylyshyn in 2001 or in *Image and Brain*. The question of whether images are depictive is taken to be resolved by imagery's activation of the topographical parts of the visual system (Kosslyn 1994, p. 405). The idea that behavioral results cannot address even basic questions about mental imagery pervades the literature; covering the imagery debates, a recent (2019) review article describes the *existence* of depictive mental representations as intrinsically inaccessible to behavioral methods:

> Originally, the debate was centered around the question of whether imagery, like perception, relies on depictive, picture-like representations or on symbolic, language-like representations. Due to imagery's inherently private nature, for a long time it was impossible to address this question. Neuroimaging studies… have now largely resolved this debate in favor of the depictive view. (Dijkstra et al. 2019)

There's nothing wrong with invoking this neuroimaging evidence for its rhetorical effect. Pointing to data indicating imagery engages perceptual mechanisms obviates the need to defend images as a theoretical construct implementable in the brain;[20] even staunch propositionalists do not doubt the existence of perception or the strength of findings in the tradition of Hubel and Wiesel. That being said, we should not confuse neuroimaging's rhetorical force for evidentiary panacea.

We may find, for instance, that "of all the brain areas that were activated during perception and during imagery, approximately two-thirds were activated in both cases"; furthermore we may notice that *Image and Brain* contains detailed theoretical developments on the basis of neuroimaging results from the occipital lobe. We may use these facts to draw parallels between imagery and perception or to claim that imagery amounts to a top-down usage of the perceptual system. In doing so, however, we should not neglect that much of the same conclusions were accessible to older methods: parallel deficits across imagery and perception (reviewed by Kosslyn (2001)), Perky's demonstration, the selective appearance of the size-disparity effect under experimental conditions implicating the imagery system, discussed above.

My claim is not that neuroimaging techniques do not produce results about cognition. They do, albeit to a far more limited extent than is often attested in the imagery literature. For example, Kosslyn describes a progression of experiments beginning with the observation that while mentally rotating inanimate objects, "the premotor cortex was activated… but in only half of the subjects." This suggested that "there could be two strategies for performing such rotations. One strategy involves imagining what you would see if you manipulated an object; the other involves imagining what you would see if someone else (or an external force, such as a motor) manipulated an object," leading to Kosslyn's discovery that inducing one or the other strategy

[20] I thank George Smith for this point.

could provoke the corresponding activation (Kosslyn et al. 2001). In demonstrating that motor involvement produced this activation, we learned something about image transformations and reaffirmed what we thought about the premotor cortex. Thus neuroimaging techniques can play a role in evidential development, though as a matter of historical fact, this role is far more limited than Kosslyn's rhetoric suggests.

The position here is thus distinct from Coltheart's important 2005 position paper denying that neuroimaging results have distinguished competing psychological theories (Coltheart 2006). From an evidentiary perspective, we've already seen several instances of neuroimaging *bearing on* theoretical distinctions, even if earlier behavioral results often supported the same conclusions. The neuroimaging results discussed in Kosslyn's 2001 paper at least evidence the existence of depictive mental representations. While one primary aim of the current paper is to correct the literature's underestimation of Kosslyn's early behavioral results, it does not deny the plurality of measures that we have thus far accrued to learn about imagery and the mind more generally. Despite the emergence of neuroimaging techniques and the new kinds of implementational knowledge they license, the evidentiary project outlined by Newell and Simon remains largely intact, or so I will argue.

3 "Neural Foundations"

Consider again the "neurocognitive" result in image rotation, above. Results like these might lead us to overestimate the physiological nature of our knowledge in imagery. This result was enabled by (1) a considerably developed computational theory of imagery characterizing image transformations, (2) the identification of a section of the cortex as typically involved in motor-related activities, and (3) a measurement of increased bloodflow to that section of the cortex under the experimental condition implicating motor imagery. The discovery of these "two strategies for performing... rotations" is a discovery about the elements of (1) responsible for implementing transformations.[21] Notice, however, that this experiment discovered very little and presupposed nothing whatsoever of the implementation of (1) such that "symbolic abstraction" needs to be abandoned. In fact, *all* of the results I have discussed thus far either develop cognitive theory or suggest large-scale features of its implementation, still always in the language of information processing.

It appears that the "neural foundations of imagery" promised in the title of Kosslyn's paper amount to specific patterns of bloodflow ("electricity," "activity," etc.) associated with some presumed cognitive functionality. The misleading practice of titling articles as such despite their content is not unique to Kosslyn. In fact, a wealth of recent papers in the cognitive science literature bear titles purporting to specify "neural foundations," "neural basis," "underpinning," etc.: see Kosslyn et al. (2001), Decety (1996), Mayseless et al. (2015), Harrington and Haaland (1999), Poeppel and Assaneo (2020), Martin and Weisberg (2003), Gold and Shadlen (2007), Rilling et al. (2002), Damasio and Geschwind (1984), to name a few. Looking to the body of these

[21] This may be somewhat understated. Changes to the transformational component of the theory may force alterations to other components.

articles, we find that the attested neural underpinning amounts to the same kinds of results we have already discussed: instances of neuroimaging developing cognitive theory or coarse sketches of its implementation, usually going no further than localization of theory-components. Neuroimaging and deficit phenomena are thus in the same evidentiary tradition.

I claim that the reason leading researchers have omitted neurophysiology from these articles is that we do not presently have a robust account of how neurons process information. Describing the true physiological underpinning of cognitive theory (or even a fragment of it) is thus simply not possible given the current state of neuroscience. Seeing as this claim is a negative existential (about what has been achieved in the study of the mind), I won't be able to prove it. We'll look to the neuroscience literature momentarily and find evidence that I am correct, but notice that the observation just made—leading researchers advertise neural implementation but never produce one—is evidence in itself of my position.[22]

To determine what has been achieved on the matter of neural information processing, let's look to perhaps the most advanced study in the physical specification of cognitive functionality: conditioning. C.R. Gallistel, a leading researcher on the topic, routinely points out that we have no physiological model of this phenomenon. Instead, we have localization of cognitive capacities, even potentially to subcellular mechanism:

> Recent electrophysiological results imply that the duration of the stimulus onset asynchrony in eyeblink conditioning is encoded by a mechanism intrinsic to the cerebellar Purkinje cell... We do not yet know the cell-intrinsic molecular mechanism that encodes the duration of the CS-US interval, nor in what part of the cellular structure it resides... That an encoding of some type occurs is, however, clear... We may make inferences about the code without knowing its physical implementation. Indeed, without such inferences we will have no idea what to look for (Gallistel 2017).

As Gallistel remarks, the long hunt for a physiological understanding of conditioning continues, guided by a constraining cognitive theory.[23] Neuroscience to this day continues its search for the engram(s), the "medium by which information extracted from past experience is transmitted to the computations that inform future behavior" (Gallistel 2021). A recent review of the subject begins with the following statement of its progress: "we do not know what an engram is fully; we have not found an engram in its entirety; we do not have a complete understanding of the biochemical and physiological parameters underlying engram storage, retrieval and updating" (Denny et al. 2017). Relatedly, a group at UCLA recently demonstrated that "RNA from sensitization-trained *Aplysia*" can "induce sensitization-like behavioral enhancement when injected into naïve recipient animals;" in other words, "RNA from a trained

[22] The (distinct) question of *why* researchers (perhaps journals) feel obliged to title their papers as such, neglecting the more accurate "possible localization of cognitive function *x*" is a sociological one. To speculate, the prefix "-neuro" and techniques from cognitive neuroscience may function honorifically in the literature. See Weisberg et al.'s (2008) "The Seductive Allure of Neuroscience Explanations" for some related discussion.

[23] Chomsky cites Gallistel to this effect; see, for instance, Berwick and Chomsky (2016, p. 50-2).

animal" produced a "learning-like behavioral change in an untrained animal." We "do not know the identity of the memory-bearing molecules at present," and thus we cannot assess their computational capacities (Bédécarrats et al. 2018).

In other words, we are yet to discover the physiological basis of memory. The implications for our discussion of imagery are significant. Most obviously, mental images are generated from long-term memory. Our ignorance of the engram thus implies physiological ignorance of at least one theory component. But the problem extends much further. Possibly the central component of Kosslyn's theory is the *visual buffer*, the "surface" onto which images are mapped (serially, from memory). The buffer enters into nearly all operations in the theory. As with long-term memory, though, this buffer itself amounts to a medium of information storage and transmission whose implementation we do not understand. Successful specification of the physical basis of imagery thus requires neurophysiological knowledge of the engram, knowledge which is presently out of reach.

Given the centrality of memory to imagery and computation in general (one of many lessons we owe to Turing's landmark (1936) paper), our ignorance of the engram forces us to adopt Newell and Simon's research strategy: model the mind in terms of its information processing capacities. Were we to reject this strategy, we wouldn't even be able to *posit* the operations in Kosslyn's theory—generation, maintenance, transformation, inspection—for these can only be described computationally. As in conditioning research, pursuit of physiological mechanism can only be entertained in the context of a cognitive theory characterizing the mechanism pursued.

From this perspective, researchers do not adopt information-processing explanations because they are committed to an arcane "computational theory of mind;" rather, computationalism construed here is a research strategy used to study mental operations despite a still-young neuropsychology, as outlined by Newell and Simon in 1972. Insofar as we lack a physiological account of cognition, this functional strategy is the only game in town capable of positing mental operations and subjecting them to sustained inquiry.

If Newell and Simon are correct about the factors motivating recourse to cognitive explanation, then many of the problems traditionally associated with cognition simply do not arise. Cognitive science would make no claims about "the scope of the mental" or the "scope of the computational." Put simply, the discipline conducts experiments as a means of adducing features of "symbolic model[s] on the basis of which pertinent specific aspects of… behavior can be calculated." While to formulate models in these terms often requires considerable abstraction from physiology, this abstraction can bear evidential fruit, as demonstrated by Kosslyn's success in *Image and Mind.* Relatedly, we need not become bogged down with questions of multiple realizability. It's plainly immaterial to the projects under consideration that Kosslyn's cognitive theory of imagery could be implemented in silicon, neurons, or a person locked in a room performing computation by hand; computationalism is a technique by which cognitive scientists have accrued evidence about the brain, evidence silent on the ontological comparability of such systems.

One *ostensible* loose end in the above account may be the frequent specification of models in terms of so-called neural nets. These amount to information processing devices capable of representing multivariate relationships, making them a useful tool

for modeling complex phenomena. One might expect given the name that we can characterize their relationship to the information processing in neurons. As with the parallel intuition for "cognitive neuroscience," however, we shouldn't trust our gut. To illustrate, consider the 2021 paper "Single cortical neurons as deep artificial neural networks," which found that "A temporally convolutional DNN [deep neural network] with five to eight layers was required to capture the I/O mapping of a realistic model of a layer 5 cortical pyramidal cell" (Beniaguev et al. 2021). As above, the subcellular mechanisms capable of implementing five-layer nets have not been specified; the authors of the article even suggest that neuroscientists "utilize this work to explore how real neurons can use their rich biophysical repertoire in order to perform specific computations from the class computed by the equivalent DNNs."

To clarify once more, I am *not* claiming that fMRI findings or successful neural net modeling are not real achievements. Localization of cognitive function is potentially important, paving the way for future more detailed work (e.g., Hubel 1982); the same holds for neural net development, as in Doris Tsao's aptly titled "The code for facial identity in the primate brain" (Chang and Tsao 2017); plainly, though, neither is "neural underpinning." As should be clear by now, that we are unable to specify neural basis does not invalidate the study of these phenomena – Kosslyn's research described in Section 1 constitutes a great evidential achievement – all it means is that we cannot neurally characterize the real and substantial discoveries made at the "functional level" (Kosslyn 1980, p. 124).

4 A Positive Role of Cognitive Neuroscience Explored

In light of the discussion above, I'd like to spend some time describing in more detail the true contributions of neuroimaging to the study of mind, looking to Kosslyn's *Image and Brain* (Kosslyn 1994). I pointed out above that both the 2001 article and *Image and Brain* successfully respond to Pylyshyn's concerns by demonstrating that imagery is implemented via top-down usage of the perceptual system. *Image and Brain* is particularly impressive in this regard because Kosslyn develops the imagery theory therein on the basis of mechanisms in high-level perception itself.

I would like to review a number of important features of this work, though a more complete treatment will have to wait for another time. To begin, the theory presented "is an extension" of the *Image and Mind* version; in most cases, "the previously inferred mechanisms have not been rejected but rather have been recast and further articulated" (Kosslyn 1994, p. 388). Accordingly, the evidence for the 1980 theory, including the experiments reviewed above, is carried over to the 1994 version. The claimed preservation of evidence across the transition from the 1980 to the 1994 theory would be an extremely interesting case study in the continuity of evidence across *prima facie* radical theory-change (Buchwald and Smith 2001).

Second, the 1994 theory remains an information processing account of mental imagery, though this is by no means apparent from Kosslyn's exposition. Consider his description of how the visual buffer changed from one theory to another:

> The visual buffer in the previous version of the theory was modeled by an array in a computer, which was anisotropic but homogeneous. The present conception is that the visual buffer corresponds to a set of topographically mapped visual areas in cortex, which are anisotropic and nonhomogeneous; resolution is greatest in the center and decreases toward the periphery. The areas [work] together to implement a multiscaled structure, and images are represented at different spatial scales within this structure. In addition, the visual buffer in the previous version of the theory was a passive receptacle of information, whereas it plays a much more active role in the present theory… In all other respects, the visual buffer plays the same role in both versions of the theory (Kosslyn 1994, p. 388-9).

If successful, the identification of a cognitive mechanism with particular brain areas has several virtues. First, neuroimaging techniques can be brought to bear on the question of involvement of that mechanism in a particular task. If one needs to determine whether a subject is using the motor strategy to implement image transformations in a particular task, one may look to activation in the premotor cortex as a kind of litmus test. Of course, there is yet no general rule for determining whether one is licensed in making inferences such as these, and a sufficiently developed cognitive account can provide principled descriptions of the circumstances under which subsystems are used without appealing to neuroimaging (Kosslyn 1994, p. 401-3). Second, learning about features of the brain region itself can inform your understanding of the cognitive functionality. For instance, it is known (by Spitzer et al. 1988) that "task demands affect the sensitivity of neurons in at least some of the areas that compose the visual buffer," suggesting that "it is possible that the resolution of the visual buffer may change, within limits, depending on task requirements" (Kosslyn 1994, p. 389). Third, we have more recently developed a limited understanding of the ways that some brain regions are interconnected (mostly in animals) by performing "electrical microstimulation combined with simultaneous fMRI" (Moeller et al. 2008). Were we able to establish in any detail the computations being performed in each of these regions, we could theoretically bring this information to bear on our cognitive understanding.

Nevertheless, exactly what is implied by identifying the visual buffer with "a set of topographically mapped visual areas in cortex" is not clear because Kosslyn never describes in any detail how its properties are implemented in these brain regions. He wants the visual buffer to implement such functions as "actively fill[ing] in missing information… and completing edges," but it remains a mystery how *any* functions, these included, are implemented in the cortex (Kosslyn 1994, p. 389). It may be that the identification of those functions and brain regions with the buffer will guide inquiry into how the cortex could accomplish such feats of computation (I hope so!). Until that is achieved, however, any account of the visual buffer will necessarily abstract away from its physiological underpinning.

The emergence of neuroimaging techniques has both deepened our functional understanding and suggested new avenues for research at the implementational level, fragments of which we have just reviewed. This research will doubtless develop considerably as the resolution of our measurements (spatial and temporal) improves. From an evidentiary perspective, the implementational research programs enabled by neuroimaging are in the same tradition as lesion studies and other defecit phenomena.

Still, though, neuroimaging research requires computational theory. At the very least, in order to characterize the mental operations under study, one needs a functional description of those operations.

Unfortunately, much contemporary imagery research ignores precise cognitive models developed on the basis of behavioral experiment, resorting instead to natural-language descriptions of the informational processes under study. To reiterate, even in imagery research, with a rich tradition of robustly developing computational models and using them to answer further questions (see Smith and Kosslyn 1980), functional theorizing beyond loose natural-language description is rare. Standard, however, are dismissive references to the tradition of research in which these models were developed without any kind of substantive engagement (e.g., Dijkstra et al. 2019, Kosslyn et al. 2001). The situation is so severe that Kosslyn himself was unable to locate a copy of any of his computational models of imagery when I asked him for one. A several-months-long communication between us, in which he attempted to contact many of his former collaborators, has thus far turned up nothing; the models could be lost. At the very least, they are not being maintained or developed by contemporary researchers. It's not exactly clear *why* the field has moved away from computational modeling: *Image and Brain* is a clear illustration of how cognitive neuroscience can mine the computational strategy profitably. One possibility is that as leading researchers herald the unique objectivity of neuroscientific methods (Dijkstra et al. 2019, Kosslyn et al. 2001), researchers trust less those theories based on behavioral experimentation. Another possibility is that researchers take neuroimaging results to be autonomous from cognitive theorizing or even to undercut the computational strategy (see Piccinini's *Neurocognitive Mechanisms* (Piccinini 2020) for some discussion of theoretical autonomy). But neither of these is defensible, again demonstrated by the successes of *Image and Brain*.

The negative evidentiary effects of this change in course are severe. Without a model constrained by the behavioral results from early in the program, the questions one asks and the answers one achieves become disconnected from this tradition of research. The "backward-looking" evidentiary process, by which early results become vindicated by later research, is made considerably less forceful. The "forward-looking" process, by which old results constrain and guide new developments (Smith and Kosslyn 1980), is largely discontinued.

5 Imagery and Syntactic Theory

If the account of the computational I have offered here is correct, then the study of visual imagery is surprisingly similar to generative grammar in its relation to the "stuff of the brain." The computational theories "have much stronger empirical support than anything available at other levels" (i.e., the brain itself), and in isolation from these theories, the neuroimaging results "are just curiosities" (Chomsky 2000, p. 24-5). What does it mean to say that "there could be two strategies" for performing mental rotations in isolation from computational theory, when our only detailed accounts of *what it is to imagine* (transform, etc.) are computational?

We have seen several positive contributions that neuroimaging techniques have made to the study of visual imagery, and we owe these to a number of ways in which this

program differs from generative syntax. First, the operations described in Kosslyn's theory are considerably more distributed across the cortex. This means that when subjects invoke specific theory components (e.g., generate and transform an image), these operations may produce distinctive patterns of activation that are observable by neuroimaging techniques. This permits the association of these operations with those brain regions, making them liable to the patterns of evidence development that I have discussed above – knowledge of the brain region can inform the cognitive account, and the cognitive account can inform our understanding of the brain region. Second, the brain-based study of the perceptual system is more advanced than the language case. Seeing as results in perception can be (and have been, see Kosslyn 1994) leveraged to learn about imagery, the cognitive neuroscience of imagery has many results to build on.

6 Conclusion

We have covered considerable ground in this essay, reaching a number of surprising conclusions.

First, we should not neglect the evidentiary successes of behavioral work, even in light of the proliferation of neuroimaging techniques. The importance of these data to the investigation of mind should not surprise us in light of Chomsky's reminder that the "study of algorithms involved in… doing long division is a study of the brain" (Chomsky 2000, p. 24). Behavioral experimentation plays a role at the cutting edge of research and, more importantly, often constitutes the evidential basis for advanced work.

Second, Kosslyn amassed considerably stronger evidence for his conclusions than he could have with any isolated experiment by *predicating active research on prior results*. When successfully employed, the epistemic payoff of this strategy is bidirectional. New explanations, usually underconstrained in psychology[24] and thus liable to being overturned, are guided by a tradition of prior results. Earlier findings, constitutively presupposed in ongoing work, passively accrue evidence. All of the imagery results I have discussed in this paper, for instance, assume a nontrivial distinction between images stored in long-term memory and those on the buffer. Psychological discoveries like these, by virtue of their entrenchment in a sustained tradition of successful research, have strong claim to permanence.

Third, despite the title of Kosslyn's 2001 article, our knowledge of imagery comprises nonphysiological computations and large-scale features of their implementation in the brain; our knowledge will continue to be computational so long as we lack a physiological understanding of how the brain stores and processes information. This fact about the limits of modern neuroscience commits cognitive scientists to Newell and Simon's research strategy: characterize the mind computationally. Neuroimaging *does* complicate the relationship between our computational account and the brain itself. These techniques may connect components of the cognitive theory to parts of

[24] Recall the "common point in the philosophy of science" to which Smith and Kosslyn referred in 1980: "empirical research in a theoretical vacuum is likely to flounder" (Smith and Kosslyn 1980).

the brain, enabling a kind of reciprocal evidential relationship between the cognitive theory and the brain region. Nevertheless, it must be stressed that in the field of imagery, neuroimaging results are meaningless outside the context of a contextualizing cognitive theory, for we cannot presently characterize imagery capacities except in the language of information processing.

If the account I have given here is correct, then surprisingly little has changed since Newell and Simon's 1972 explanation for their adopting an information processing account of human problem solving – this remains the only way to model high-level mental phenomena in detail. We may be in what Jackendoff calls the "Age of Cognitive Neuroscience" by virtue of the incredible emphasis placed on neuroimaging techniques in the literature (Jackendoff 2002);[25] cognitive neuroscience may even have ushered in a kind of *conceptual* revolution concerning the relationship between cognition and the brain — some researchers doubtless did conceive of cognition as autonomous in principle from neuroscience/neuroimaging (Piccinini 2020). That being said, the evidentiary project outlined by Newell and Simon remains intact, no more undercut by cognitive neuroscience than lesion data.[26]

References

Azzouni, Jody. 2020. *Attributing Knowledge: What it Means to Know Something*. Oxford University Press.

Bédécarrats, Alexis, et al. 2018. RNA from trained Aplysia can induce an epigenetic engram for long-term sensitization in untrained Aplysia. *Eneuro* 5(3).

Beniaguev, David, Idan Segev, and Michael London. 2021. Single cortical neurons as deep artificial neural networks. *Neuron* 109 (17): 2727–2739.

Berwick, Robert C., and Noam Chomsky. 2016. *Why only us: Language and evolution*. MIT press.

Buchwald, Jed Z., and George E. Smith. 2001. Incommensurability and the discontinuity of evidence. *Perspectives on Science* 9 (4): 463–498.

Chang, Le., and Doris Y. Tsao. 2017. The code for facial identity in the primate brain. *Cell* 169 (6): 1013–1028.

Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Christiansen, Morten H., and Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31 (5): 489–509.

Christiansen, Morten H., and Nick Chater. 2009. The myth of language universals and the myth of universal grammar. *Behavioral and Brain Sciences* 32 (5): 452–453.

Coltheart, Max. 2006. What has functional neuroimaging told us about the mind (so far)? In *European Cognitive Neuropsychology Workshop, 2005, Bressanone, Italy; Position paper presented to the aforementioned conference*. Masson Italia.

Damasio, Antonio R., and Norman Geschwind. 1984. The neural basis of language. *Annual Review Of Neuroscience* 7 (1): 127–147.

Decety, Jean. 1996. The neurophysiological basis of motor imagery. *Behavioural Brain Research* 77 (1–2): 45–52.

Denny, Christine A., Evan Lebois, and Steve Ramirez. 2017. From engrams to pathologies of the brain. *Frontiers in Neural Circuits* 11: 23.

Dijkstra, Nadine, Sander E. Bosch, and Marcel A.J.. van Gerven. 2019. Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences* 23 (5): 423–434.

Gallistel, Charles R. 2017. The coding question. *Trends in Cognitive Sciences* 21 (7): 498–508.

[25] It may be that researchers emphasize as they do because they suspect neuroimaging techniques will eventually unveil in great detail the workings of the mind, but this is a paper about evidence-based achievements in the present.

[26] See Buchwald and Smith (2001) for a discussion of the difference between conceptual and evidentiary revolutions.

Gallistel, Charles R. 2021. The physical basis of memory. *Cognition* 213: 104533.
Gold, Joshua I., and Michael N. Shadlen. 2007. The neural basis of decision making. *Annual Review of Neuroscience* 30: 535–574.
Harrington, Deborah L., and Kathleen Y. Haaland. 1999. Neural underpinnings of temporal processing: A review of focal lesion, pharmacological, and functional imaging research. *Reviews in the Neurosciences* 10 (2): 91–116.
Hubel, David H. 1982. Exploration of the primary visual cortex, 1955–78. *Nature* 299 (5883): 515–524.
Huettig, Falk. 2015. Four central questions about prediction in language processing. *Brain Research* 1626: 118–135.
Jackendoff, Ray S. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. USA: Oxford University Press.
Kosslyn, Stephen M., et al. 1977. Category and continuum in mental comparisons. *Journal of Experimental Psychology: General* 106 (4): 341–375.
Kosslyn, Stephen M. 1980. *Image and Mind*. Harvard University Press.
Kosslyn, Stephen M., et al. 1983. Generating visual images: Units and relations. *Journal of Experimental Psychology: General* 112 (2): 278–303.
Kosslyn, Stephen M., et al. 1984. Individual differences in mental imagery ability: A computational analysis. *Cognition* 18 (1–3): 195–243.
Kosslyn, Stephen M. 1994. *Image and brain: The resolution of the imagery debate*. MIT press.
Kosslyn, Stephen M., et al. 2001. Imagining rotation by endogenous versus exogenous forces: Distinct neural mechanisms. *NeuroReport* 12 (11): 2519–2525.
Kosslyn, Stephen M., Giorgio Ganis, and William L. Thompson. 2001. Neural foundations of imagery. *Nature Reviews Neuroscience* 2: 635–642.
Martin, Alex, and Jill Weisberg. 2003. Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology* 20 (3–6): 575–587.
Mayseless, Naama, Ayelet Eran, and Simone G. Shamay-Tsoory. 2015. Generating original ideas: The neural underpinning of originality. *Neuroimage* 116: 232–239.
Miller, Brett, Neil Myler, and Bert Vaux. 2016. Phonology in universal grammar. In *The Oxford Handbook of Universal Grammar*.
Moeller, Sebastian, Winrich A. Freiwald, and Doris Y. Tsao. 2008. Patches with links: A unified system for processing faces in the macaque temporal lobe. *Science* 320 (5881): 1355–1359.
Newell, Allen, and Herbert A. Simon. 1972/2019. *Human Problem Solving*. Echo Point Books and Media. isbn: 978-1-63561-792-4.
Paivio, Allan. 1975. Perceptual comparisons through the mind's eye. *Memory & Cognition* 3 (6): 635–647.
Perky, Cheves West. 1910. An experimental study of imagination. *The American Journal of Psychology* 21 (3): 422–452.
Piccinini, Gualtiero, and Andrea Scarantino. 2010. Computation vs. information processing: Why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A* 41 (3): 237–246.
Piccinini, Gualtiero. 2020. *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford University Press.
Poeppel, David, and M. Florencia Assaneo. 2020. Speech rhythms and their neural foundations. *Nature Reviews Neuroscience* 21 (6): 322–334.
Pylyshyn, Zenon W. 1984. *Computation and Cognition*. MIT press.
Rilling, James K., et al. 2002. A neural basis for social cooperation. *Neuron* 35 (2): 395–405.
Shrout, Patrick E., and Joseph L. Rodgers. 2018. Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology* 69: 487–510.
Smith, George E., and Stephen M. Kosslyn. 1980. An information-processing theory of mental imagery: A case study in the new mentalistic psychology. In PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association. Vol. 1980. 2. Philosophy of Science Association, 247–266.
Smith, George E. 2010. Revisiting accepted science: The indispensability of the history of science. *The Monist* 93 (4): 545–579.
Smith, George E. 2014. Closing the loop. In *Newton and Empiricism*, 262–352.
Smith, George E., and Raghav Seth. 2020. *Brownian Motion and Molecular Reality: A Study in Theory-Mediated Measurement*. Oxford University Press.
Spitzer, Hedva, Robert Desimone, and Jeffrey Moran. 1988. Increased attention enhances both behavioral and neuronal performance. *Science* 240 (4850): 338–340.

Turing, Alan Mathison. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 2: 230–265.
Wagenmakers, Eric-Jan. 2007. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review* 14 (5): 779–804.
Weisberg, Deena Skolnick, et al. 2008. The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience* 20: 470–477.
Wiesel, Torsten N. 1982. Postnatal development of the visual cortex and the influence of environment. *Nature* 299 (5884): 583–591.