# Object files encode possible object identities, but not possible locations

**Gabor Brody (gaborbrody@protonmail.com)**
Department of Cognitive, Linguistic & Psychological Sciences, Brown University, 190 Thayer Street
Providence, RI 02912 USA

**Peter Mazalik (pmazali1@jhu.edu)**
Department of Psychological and Brain Sciences, The Johns Hopkins University, 3400 N. Charles Street
Baltimore, MD 21218-2686 USA

**Roman Feiman (roman_feiman@brown.edu)**
Department of Cognitive, Linguistic & Psychological Sciences, Brown University, 190 Thayer Street
Providence, RI 02912 USA

## Abstract

It is uncontroversial that humans can represent possibilities, but it is debated what this claim amounts to. Under broad views of modal cognition, many representational and reasoning systems represent possibilities at multiple levels of cognitive architecture. Under narrow views of modal cognition, there exists a special kind of higher-level modal thought, that can be measured with purpose built non-verbal modal cognition tasks. Here we ask whether object tracking mechanisms that are assumed to lack the higher-level narrow modal capacity, show behavioral signatures that are assumed to require it. We find signature of modal representation in one task, but not another. The finding suggests that there is no clear difference between tasks that tap broad and narrow modal cognition, and invites a reassessment of the evidence for the latter.

**Keywords:** modal cognition; object representation; object files; possibility; reasoning; perception

## Introduction

Agents have to represent the world not only as it *is* but also as it *could be.* Representations that distinguish what is actual from what is possible belong to the domain of modal cognition. Modal cognition so broadly construed is omnipresent in cognitive life (Friston, 2010; Griffiths & Tenenbaum, 2006; Guan & Firestone, 2020; Lewis, 1986; Phillips & Kratzer, 2022). For instance, decision making and action planning are generally modeled as processes that generate and select from multiple possible, but yet-to-be actualized decisions and actions (Kording, 2007; Trommershäuser et al., 2008). Perception – especially according to theories that involve hypotheses and predictions – applies information from the senses to representations and models of how the world could be with the goal of achieving an accurate model of the actual world (Helmholtz, 1948; Knill & Richards, 1996; Weiss et al., 2002). Understanding natural language also requires entertaining possible ways the world could be. Even simple declarative sentences are argued to have meanings that specify the possible cases in which they are true, while questions are argued to have meanings that specify the possible declarative sentences that are good answers to them. (Barwise & Perry, 1981; Karttunen, 1977; Kratzer, 1981; Lewis, 1972; Stalnaker, 1984). If such meanings are represented in any way, then understanding a sentence requires representing the relevant possibilities. Under this broad notion of modal cognition, the difference between representing possibilities in, say, perception and sentence comprehension could be cashed out not in terms of their capacity to represent possibilities, but in terms of other inherent differences between how the two system operate.

According to some psychological theories, beyond such a broad modal capacity, there is also a special high-level *narrow* capacity to represent possibilities. This capacity is argued to afford further representational and inferential abilities that provide humans with powerful, possibly species-specific and ontogenetically later-emerging reasoning capacities (Carey et al., 2020). What makes narrow possibility representations special differs between views, but is generally argued to involve specific *formal* properties of the represented possibilities, which might imbue them with unique *functions*. One such claim is that only these narrow possibility representations are encoded with explicit, composable, and syntactically detachable mental symbols for modal notions, like possibility and necessity (Leahy & Carey, 2020). Another claim is that only the narrow possibility representations encode categorical distinctions between impossibility and possibility, or between possibility and necessity, while broad modal cognition can only encode graded, probabilistic distinctions (Cesana-Arlotti et al., 2020; Cesana-Arlotti et al., 2012; Rips, 2001).

Good evidence for a psychologically distinct, narrow modal capacity would be a distinctive behavioral competence. If there is a set of tasks that require a single underlying psychological capacity to perform successfully, then having this capacity could enable success on all these tasks, while not having it should predict common failure. Critically, this variation in performance should also be independent of the more ubiquitous processes that fall under broad modal cognition. That is, there should be some things that creatures lacking the narrow modal capacity cannot do, even if they have other highly sophisticated abilities.

There is currently wide agreement about what kinds of tasks can diagnose narrow modal cognition. These mostly come in two main flavors (c.f. Goddu et al., 2021):

(1) *Possible identity tasks*. Participants have to encode that there are multiple ways objects can be mapped to identities. For instance, in a study by Cesana-Arlotti et al. (2022), participants see a partially occluded object, and cannot tell whether it is a flower or a dinosaur.



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

(2) *Possible location tasks*. Participants have to encode multiple possible trajectories or locations for a single object. In the paradigm study by Redshaw and Suddendorf (2016), a ball is dropped into an inverted Y-shaped tube, so participants cannot tell whether it will come out on the left or the right.

In both kinds of tasks, participants' behavior is used to establish whether they have encoded the relevant possibilities. This can be a direct behavioral measure (e.g., whether they cover both possible openings of the Y-shaped tube) or a more indirect measure (e.g., pupil dilation as a proxy for increased cognitive effort used to represent multiple possibilities). Note that although indirect measurements like pupil-dilation might seem to measure perceptual processes, they are argued to measure low-level correlates of high-level reasoning.

A mixed pattern of successes and failure across tasks has produced a debate over which creatures can be credited with narrow modal cognition. According to *continuity* views, the evidence is converging to show that young infants and/or other species succeed on tasks that tap narrow modal cognition, while the failures can be attributed to performance demands (Alderete & Xu, 2023; Cesana-Arlotti et al., 2020; Cesana-Arlotti et al., 2018; Cesana-Arlotti et al., 2022; Engelmann et al., 2023; Engelmann et al., 2021; Goddu et al., 2021; Phillips & Kratzer, 2022; Turan-Küçük & Kibbe, 2024). On *discontinuity* views, the specific tasks that infants or other species pass could be solved without a full command of narrow modal cognition; the real diagnostic convergence is between tasks on which success begins in the preschool years (Carey et al., 2020; Leahy, 2023; Leahy et al., 2022; Leahy & Carey, 2020; Redshaw & Suddendorf, 2016).

Nevertheless, both continuity and discontinuity views share two key assumptions: that there is a distinct narrow capacity, and that it can be measured with tasks that require representing the possible locations and identities of objects.

In this paper, we want to stress-test these shared assumptions. We ask whether tasks that require representing possible identities and locations can be solved by mechanisms that are part of low and mid-level perceptual and attentional processing, which cannot have modal representations in the narrow sense. We focus on so called 'object files', which keep track of objects' locations and identities to provide information about 'what goes where'. Object files encode separate objects as individuals and maintain them through motion, transformation, and even occlusion (Kahneman et al., 1992; Pylyshyn, 2007).

As perceptual mechanisms, object files meet none of the criteria for narrow modal thought: they are not high-level (in particular, they do not require language), they are not systems for reasoning, and they are well known to be neither species-specific nor late developing in humans (Carey, 2000; Leslie et al., 1998; Scholl & Leslie, 1999). At the same time, object files might be one of the few representational systems where there are reasons to doubt modal processes are present even in the broadest sense. It has been argued that the tracking mechanism for object files is non-attentive (Alvarez et al., 2005), bottom-up (Haladjian & Pylyshyn, 2006) and consists of heuristic rather than predictive processing (Keane & Pylyshyn, 2006).

Therefore, if object files can be used to represent information that is supposed to be the sole province of narrow modal cognition, it would challenge standard views of the narrow capacity, and of whether existing methods test it. On the other hand, if perceptual mechanisms do fail on these tasks, it would provide further converging evidence that these tasks do what they are supposed to do and measure narrow modal thought (see Jasbi et al., 2019).

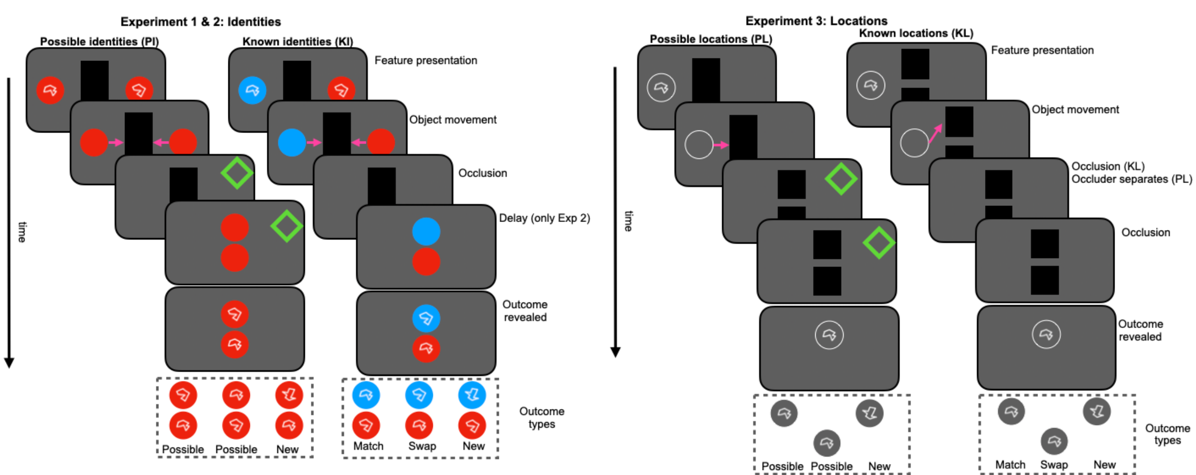


Figure 1: Trial structure of the experiments. In Experiment 1 & 2 shapes are revealed on discs that are either the same color (Possible identities) or different colors (Known identities). The shapes disappear and the discs move behind occlusion. Then the occluder is removed revealing one of the outcome types. In Experiment 3 one disc either moved behind a large occluder that later separates (Possible locations), or behind one of two occluders (Known locations). Green diamonds mark periods where representing two mutually exclusive possibilities could prime participants' responses to *both* possible outcomes.

In order to evaluate whether object files can represent possibilities, we rely on the OSPB (object specific preview benefit) phenomenon: when a feature is briefly flashed on an object, participants are faster to notice it reappearing later on the same object than on another object, *even* if the objects moved, got occluded, or changed in other ways (Hollingworth & Franconeri, 2009; Kahneman et al., 1992). OSPB is generated because features get bound to visual objects, which primes processing of the same features specifically when they reappear on the same visual objects. It is precisely because OSPB dissociates from higher level thought (Mitroff et al., 2005) that it is well-suited to testing whether tasks that require representing possible object identities and locations tap into narrow modal cognition.

## Experiment 1: Possible identities

We aimed to establish whether multiple possible object identities can be encoded in the object file system using the OSPB effect. We test this by measuring OSPB to events that invoke two *equally possible* ways features can be mapped to objects. If there is an OSPB for the merely possible identities of an object that is on par with deterministic identities, that would be evidence that two objects are represented as *possibly being the same* as one previously seen object, all within the object file system. This would imply that object files encode possibilities. On the other hand, if participants do not generate OSPB for the possible events – and especially if these are treated on par with incongruent events, on which the previewed feature mismatches the subsequent display – that implies that possibilities are not being represented.

### Methods

**Participants** We tested 45 participants on Prolific (English fluent, from the US and UK) who received $7 compensation for partaking in the experiment which took around 25-30 minutes. Five of these participants were excluded for lower than 65% correct responses (N=39). No other demographic data was collected. Participants gave informed consent. The study was approved by the Brown IRB (#1808002199).

**Design & Stimulus** Our materials and setup were based on the feature-based OSPB paradigm of Hollingworth and Franconeri (2009). Two distinct shapes were briefly presented on two colored discs (Figure 1). The discs were then put in motion and got occluded at the mid-point of a single central occluder. After occlusion, the discs were revealed again, with the shapes visible. Participants were tasked with deciding whether or not these final shapes were the same shapes as the initial ones, by pressing buttons on their keyboard. They were told to disregard any other information. Critically, depending on the condition, the two discs either differed in color or shared the same color. When the discs differed in color, their identities after reappearance were known (*KI*). Our aim here was to replicate the basic OSPB finding and provide a baseline. When the discs shared the same color, participants could only represent *possible* identities (*PI*): post-occlusion, each disc could be either of the initial ones. *PI* is thus our critical manipulation. We presented 3 outcome types following *KI* setups: *new* (one of the shapes changed), *match* (same shapes, congruent discs), and *swap* (same shapes, swapped discs), and 2 outcome types for *PI*: *new* (one shape is new), and *possible* (same shapes, but could be mapped onto the discs in two possible ways). Participants saw each of the 5 different trial types presented 51 times, resulting in 255 trials/participant. Our main question was how quickly participants responded on *PI-possible* trials: on par with *KI-swap* trials (where no OSPB is expected), or on par with *KI-match* trials (where there should be an OSPB)?

### Results

**OSPB Baseline** We started by removing errors and outliers (>4000ms; <500). We then established the OSPB in the *KI* condition by fitting a model to the *KI* data, in which we Helmert-coded the trial outcome (`RT ~ outcome + (1|participant)`). We found an OSPB of 33ms (*match* vs *swap*: β=21.46, *p*=.002), while *match* & *swap* together showed faster responses than *new* trials (β=30.53, *p*<.001). See Figure 2 for data.

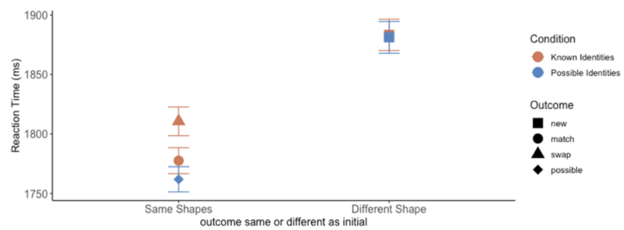


Figure 2: Mean reaction time across conditions (*KI/PI*) and outcomes in Experiment 1. Error bars represent 1 SE

**OSPB for possible outcomes** We proceeded to fit two 2x2 models (`RT ~ condition * outcome + (1|participant)`). The models differed in which two of the three *KI* outcomes were included (first model: *match***,** *new*; second model: *swap***,** *new*), but we kept the *PI* outcomes constant (*possible*, *new*), so that *PI-possible* was grouped into one variable with *KI-match* in the first model and with *KI-swap* in the second. The first model revealed only a main effect of outcome (*match*/*possible* faster than *new*: β=103.25, *p*<.001), and no interaction (β=18.54, *p*=.3). Thus, we found no evidence that the *match* vs. *new* distinction works differently than the *possible* vs. *new* distinction. In the second model (with *KI-swap* data) we found both a main effect of outcome (*swap/possible* faster than *new*: β=103.25, *p*<.001) and of condition (*PI* faster than *KI*: β=50.87, *p*=.001). Crucially, we also found an interaction due to *PI-possible* trials showing lower RTs (i.e., a greater OSPB) compared to *KI-swap* trials (β=49.61, *p*=.026).

### Discussion

We found three effects of interest. First, we validated the paradigm and created a baseline by showing that when possibilities do not need to be represented, participants generate an OSPB. Second, we found that in trials which raised two possibilities for which feature maps to which object file, both of those possibilities were primed. Third, we found that trials that involved possibilities differed from unprimed trials, in which the features that reappeared were different than those at the start. Taken together, this set of findings is most straightforwardly explained if both possible ways that object identities could be mapped onto the visual objects (i.e., both possibilities for which object is which) were primed. One limitation of this finding is that the shapes and the colors re-appeared at the same time. This makes it possible that the shapes were re-identified before the colors. While this alternative explanation alone would not explain why the different color swap trials resulted in a slowdown, it would still be good to know whether participants can maintain possible identity-objects mappings for a longer period. This would also be a closer match to prior possible identities tasks (e.g. Cesana-Arlotti et al., 2022).

## Experiment 2: possible identities and delay

We aimed to replicate and extend Experiment 1 by adding a 1000ms delay between the reappearance of the discs and the reappearance of the shapes on top of them. This meant that participants had to entertain possible mappings between the features and the objects for longer than in Experiment 1.

### Methods

The methods, design, and stimuli of this experiment were exactly as in Experiment 1, except that we added a 1000ms delay between when the occluder disappeared to reveal the discs, and when the identifying shapes subsequently reappeared on top of the discs.

**Participants** We tested 45 new participants of which 7 were eliminated for low performance (final N=38).

### Results

The analysis plan was identical to Experiment 1.

**OSPB Baseline** After removing outliers and errors, we started by establishing the OSPB in the *KI* condition. We found an OSPB of 69 ms (*match* vs *swap*: β=30.78, $p<.001$). Again, we found that *match*&*swap* together was faster than *new* (β=19.79, $p<.001$).

**OSPB for possible outcomes** We to fit two 2x2 models to data that both included both *PI* outcomes (*possible, new*), but differed in which two *KI* outcomes they included (first model: *match*, *new*; second model: *swap*, *new*). The first model revealed only a main effect of outcome (*match*/*possible* faster than *new*: β=103.12, $p<.001$), and no interaction (β=-23.24, $p=.283$). In the second model (with *KI-swap* data), we found both a marginal effect of outcome (*swap*/*possible* faster than *new*: β=29.81, $p=.071$) and a significant effect of condition (*PI* faster than *KI*: β=56.84, $p<.001$). Crucially, we also found an interaction due to *PI-possible* trials being primed more than *KI-swap* trials (β=49.43, $p=.028$). Taken together, just as in Experiment 1, we found that *PI-possible* trials are treated differently from *KI-swap* trials, but not from *KI-match* trials.

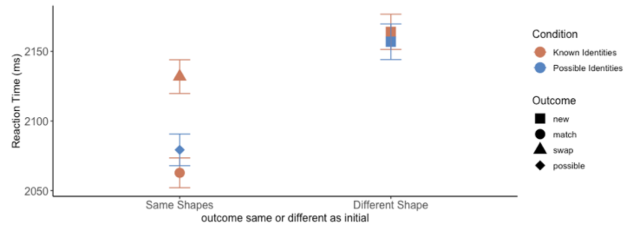


Figure 3: Mean reaction time across conditions (*KI/PI*) and outcomes in Experiment 1. Error bars represent 1 SE.

### Discussion

Even with one second of delay, during which participants had to maintain possible mappings between discs and shape features, Experiment 2 replicated all the relevant effects from Experiment 1. Participants generated an OSPB when the trial was deterministic, which was comparable to when the trials involving possibilities implied that that they represented those possibilities. This implies that possible identities can be maintained for longer timeframes in line with how object files can be maintained for a longer period (Noles et al., 2005).

## Experiment 3: possible locations

Experiment 3 changed the nature of the available possibilities. Instead of representing different possible identities for two different objects, in this task, participants saw one object disappear behind an occluder, such that it could reappear in one of two possible locations.

### Methods

**Participants** We tested 45 new participants of which 3 were excluded for low performance (N=42).

**Design & Stimulus** The structure of Experiment 3 was the same as of Experiment 1, with the following change. In every trial, there was only a single disc. Depending on the trial, the disc either moved behind one of two spatiotemporally distinct occluders, such that it had a known location (*KL*), or else it moved behind one larger occluder that later separated into two distinct occluders, so that the disc could be behind either, in one of two possible locations (*PL*). These two trial types are the location equivalent to *KI* and *PI* trials from the previous experiments. The outcomes had the same structure as before (match, swap, new, and possible), where for instance a *KL-swap* trial involved revealing the correct shape, but the disc was not behind the correct occluder (Figure 1).

### Results

The analysis plan was exactly as before, but the `condition` variable encodes location (*KL*/*PL*) instead of identity.

**OSPB** After cleaning outliers and errors, we began by establishing the OSPB, now in the *KL* condition. Again, we found an OSPB of 62ms (*match* vs *swap*: β=31.46, $p=.002$). *match*&*swap* together were also faster than *new* (β=22.93, $p<.001$).

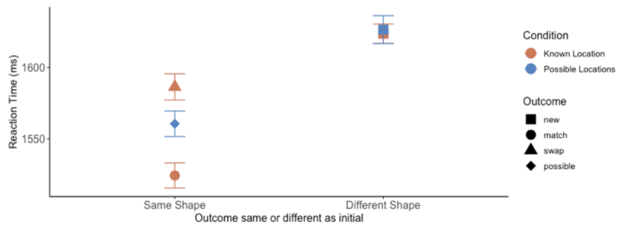


Figure 5: Mean reaction time across conditions (*KL/SL*) and outcomes in Experiment 3. Error bars represent 1 SE.

**OSPB for possible outcomes** Just as before, we fit two 2x2 models (`RT ~ condition * outcome + (1|participant`). The models differed in which two of the three *KL* outcomes were included (first model: *match*, *new*; second model: *swap*, *new*) for comparison with the same *PL* data (*possible*, *new*). The first model revealed a main effect of outcome (*match*/*possible* faster than *new*: β=99.59, *p*<.001), a main effect of condition (*KL* faster than *PL*: β=36.09, *p*=.003) and a significant interaction (β=99.59, *p*=.039). These effects were all driven by participants not responding as fast on the *PL-possible* trials compared to the *KL-match* trials. In the second model (with *KL-swap* data), we found a main effect of outcome (*swap/possible* faster than *new*: β=37.53, *p*<.01) and a main effect of condition (*KL* faster than *PL*: β=25.8, *p*=.034). When looking at the interaction between these two terms, we found a statistical tendency (β= 28.68, *p*=.075). Taken together, we found evidence showing that *PL-possible* trials were not encoded the same way as *KL-match* trials, and found some indication that they were also encoded differently from *KL-swap* trials.

### Discussion

Experiment 3 replicated the basic OSPB effect when the location of an object could be unambiguously reidentified, but found no evidence that possible locations were represented. Participants were slower to respond on possibility trials compared to when they knew where the objects were. The evidence implies that the object files participants used to track an object behind occlusion did not encode multiple possible locations for where it could end up. On the other hand, the data could be explained by positing that instead of representing possibilities, the object file system *guessed* which occluded location the object is at on *PL* trials. After all, response times for *possible* trials sit in between swap and match trials, which would make sense if participants were guessing, effectively encoding half of the *possible* outcome trials as *match* and the other half as *swap*. To be clear, this interpretation is highly speculative; we come back to it in the general discussion.

## General Discussion

We found that one particular representational system, object files, was able to represent the possible ways two object identities can be matched to two visual objects, but was unable to represent two possible locations for a single object. The successes we found are very surprising if we assume that this type of task can only be solved by high-level narrow modal representations (Cesana-Arlotti et al., 2022), which might be species-specific or late-developing (Carey et al., 2020). This finding challenges how the distinction between narrow and broad modal cognition is construed, at least in terms of how the narrow capacity is operationalized. At the same time, this finding does not show that any of the previous tasks were solved by broad processes. Rather, it implies that even if the distinction is maintained, the relevant studies may not specifically tap narrow modal cognition, and therefore may not be able to adjudicate between continuity and discontinuity theories. Before exploring this conclusion, we must first address alternative explanations of our findings.

### Alternative explanations

Is it possible that the possible identities task might not have measured OSPB, a perceptual effect, but instead accidentally measured narrow modal reasoning? In this case, the effects we find would be top-down influences from this narrow capacity on perception, rather than effects within perception itself. There are two main reasons to doubt this interpretation. First, we decided to test mid-level perceptual processing and the OSPB specifically because it has been found to be very specific to object file processing. It is so specific, that it even dissociates from conscious *perceptual* experience (Mitroff et al., 2005). The other reason is the contrast with the locations task: it is unclear why participants would reason explicitly in only one of the two logically equivalent tasks.

Another way to contest the successes in the possible identities tasks is to challenge a critical assumption in our design: that the trials that raise different possibilities operate with the same object file representations as the deterministic trials. For instance, it is possible that instead of measuring a *priming benefit* in the possibility trials, we actually measured *incongruence detection*. For instance, participants might only encode the shapes separately, without binding them to visual objects. This would still let them notice the new shapes, without having to encode possible ways to map shapes to identities. While not impossible, this alternative fails to explain why object files would function this way specifically in the possibility-involving but not in the deterministic trials (where the *swap* trials being slower than *match* trials is evidence for active binding). Another reason to doubt this interpretation is again the contrast with the location task. If there are perceptual non-possibility-involving strategies available to solve the identities task, it is unclear why they were not employed in the locations task.

In sum, although we cannot definitively rule out these alternative explanations, they require significant revisions to established accounts of object perception and the OSPB phenomenon. Critically, they struggle to explain the differential patterns between the location and identities tasks.

### Locations and identities in object files

As we discussed spatial tracking of object files has been argued to be “non-inferential” (Franconeri et al., 2012; Keane

& Pylyshyn, 2006), "bottom-up" (Haladjian & Pylyshyn, 2006) and "non attentive" (Alvarez et al., 2005). These arguments fit neatly with the results of Experiment 3, where we found no evidence that this tracking system is able to entertain multiple possible locations. However, object files are argued to be non-inferential in how they track objects, not in how they bind *identity* information to them. Indeed, tracking mechanisms and feature binding are argued to work differently in the object file system (Pylyshyn, 2007; Scholl, 2001). Experiments 1 & 2 extend our understanding of this system, showing that object files are also capable of actively representing multiple possible object-to-feature mappings.

## Continuity vs discontinuity in modal cognition

Recall that both continuity and discontinuity views assume that there is a distinct, narrow, high-level capacity for modal thought, and that tasks that require representing either multiple possible identities or locations for an object measure this capacity. What they differ in is their evaluation of the empirical landscape. Continuity theorists believe that there is sufficient evidence showing that infants and other species have command of the narrow capacity, while discontinuity theorists believe that the pattern of failures is robust enough that explaining away the early/cross-species successes is more economical.

On first blush, our results seem to provide ammunition for the discontinuity theorists. If perceptual mechanisms can represent possible identities, as we show, then other studies where infants or other species succeed in representing multiple possible identities (Cesana-Arlotti et al., 2020; Cesana-Arlotti et al., 2018; Cesana-Arlotti et al., 2022) can be explained by citing these very mechanisms (Feiman et al., 2022; Jasbi et al., 2019). If perceptual mechanisms cannot represent possible *locations*, all the more reason to think that the possible-location tasks on which 3-year-olds fail (Leahy, 2023; Leahy et al., 2022; Leahy & Carey, 2020) might require a separate, later-developing capacity.

But on a closer look, continuity theorists can also find support in our results. First, possible-location and possible-identity tasks are structurally similar, differing only in *which* possibilities participants need to represent. If perceptual mechanisms can solve structurally similar tasks to ones where three-year-old children fail, it is all the more likely that those failures implicate increased performance demands. Second, evidence from the location tasks on which younger children fail is compatible with the interpretation that the object file system is, in effect, guessing one of the two possible locations. If this is the case, then any high-level competencies in location tasks could be masked by this perceptual guessing process, which would generate an unreliable and (half of the time) incorrect representation of where the object is. In this way, these studies can also provide possible ways to explain away three-year-olds' failures on possible-location tasks, no matter what explains younger children's successes on possible-identity tasks.

## Modal cognition broad and narrow

Multiple proposals for what makes narrow high-level modal cognition special are explicated in terms of syntactic or logical *form*. But empirically, we lack direct ways to probe formal properties of thoughts. This is why modal cognition tasks look at *function* – increased inferential or representational capacity – in order to establish these formal claims. The main assumption is that passing certain modal cognition tasks *requires* having the specific formal apparatus. Our findings pose a challenge to this assumption. We looked at a part of the mind that should lack this formal apparatus and found that it passes one of the two tasks. We discuss three ways this assumption can be revised in light of our findings.

One option is to give up on these tasks as true measures of narrow modal thought, but hold on to the rest of the characterization of the capacity. Maybe there is such a capacity and having it does have unique functional consequences, but these tasks just do not tap them. The main issue with this option is that it gives up on much of the empirical support for the narrow notion in general, and for discontinuity theories in particular. One still wants a precise explanation of what the capacity is that 3-year-old children lack, causing them to fail on certain tasks. This highlights a desperate need for linking hypotheses between narrow modal cognition and its empirical measurements. What is it about the original possible identity or location tasks that is supposed to engage narrow modal thought, and why those tasks in particular? More generally, what abilities are afforded only to those who have the narrow capacity, and why those?

The second option is to give up on the assumption that narrow modal thought affords unique types of behavioral capacities, but still hold on to the idea that it has unique formal properties. One natural revision is to posit that the unique form of narrow modal thought derives from the fact that it is part of the natural language capacity. While this option could not be tested by success or failure on any special set of tasks, it might be possible to target by other means, such as structural priming of modal meanings.

The third option is to explicitly reject that there are special narrow ways to represent possibilities but hold on to the tasks as good measures of modal cognition broadly. This view provides a simple intuitive notion of what it means to be a possibility representation (a representation of a possibility) while still explaining differences in modal thought by appealing to differences between the systems that compute over possibilities. We already know that natural language and object files are different cognitive systems: natural language has proprietary syntactic and semantic structure. If these systems already work differently, that might suffice to explain why possibilities behave differently within them. We see a unified view of possibility representations – as a basic currency of human thought – as one possibility worth pursuing, at least for methodological reasons. Only if broad modal cognition proves inadequate to explain observed phenomena, is it worth positing an additional narrow capacity.

## References


Alderete, S., & Xu, F. (2023). Three-year-old children's reasoning about possibilities. *Cognition*, *237*, 105472.

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 643.

Barwise, J., & Perry, J. (1981). Situations and attitudes. *The Journal of Philosophy*, *78*(11), 668-691.

Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development*, *1*(1), 37-41.

Carey, S., Leahy, B., Redshaw, J., & Suddendorf, T. (2020). Could it be so? the cognitive science of possibility. *Trends in Cognitive Sciences*, *24*(1), 3-4.

Cesana-Arlotti, N., Kovács, Á. M., & Téglás, E. (2020). Infants recruit logic to learn about the social world. *Nature communications*, *11*(1), 5999.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, *359*(6381), 1263-1266.

Cesana-Arlotti, N., Téglás, E., & Bonatti, L. L. (2012). The probable and the possible at 12 months: Intuitive reasoning about the uncertain future. *Advances in child development and behavior*, *43*, 1-25.

Cesana-Arlotti, N., Varga, B., & Téglás, E. (2022). The pupillometry of the possible: an investigation of infants' representation of alternative possibilities. *Philosophical Transactions of the Royal Society B*, *377*(1866), 20210343.

Engelmann, J. M., Völter, C. J., Goddu, M. K., Call, J., Rakoczy, H., & Herrmann, E. (2023). Chimpanzees prepare for alternative possible outcomes. *Biology Letters*, *19*(6), 20230179.

Engelmann, J. M., Völter, C. J., O'Madagain, C., Proft, M., Haun, D. B., Rakoczy, H., & Herrmann, E. (2021). Chimpanzees consider alternative possibilities. *Current Biology*, *31*(20), R1377-R1378.

Feiman, R., Mody, S., & Carey, S. (2022). The development of reasoning by exclusion in infancy. *Cognitive Psychology*, *135*, 101473.

Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, & Psychophysics*, *74*, 691-702.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, *11*(2), 127-138.

Goddu, M. K., Sullivan, J. N., & Walker, C. M. (2021). Toddlers learn and flexibly apply multiple possibilities. *Child development*, *92*(6), 2244-2251.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767-773.

Guan, C., & Firestone, C. (2020). Seeing what's possible: Disconnected visual parts are confused for their potential wholes. *Journal of experimental psychology: general*, *149*(3), 590.

Haladjian, H. H., & Pylyshyn, Z. W. (2006). Implicit multiple object tracking without an explicit tracking task. *representations*, *67*(2), 324-334.

Helmholtz, H. v. (1948). Concerning the perceptions in general, 1867.

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition*, *113*(2), 150-166.

Jasbi, M., Bohn, M., Long, B., Fourtassi, A., Barner, D., & Frank, M. C. (2019). Comment on Cesana-Arlotti et al.(2018).

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175-219.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*, 3-44.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology*, *52*(4), 346-368.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

Kording, K. (2007). Decision theory: what" should" the nervous system do? *Science*, *318*(5850), 606-610.

Kratzer, A. (1981). The notional category of modality. *Formal semantics: The essential readings*, 289-323.

Leahy, B. (2023). Don't you see the possibilities? Young preschoolers may lack possibility concepts. *Developmental Science*, e13400.

Leahy, B., Huemer, M., Steele, M., Alderete, S., & Carey, S. (2022). Minimal representations of possibility at age 3. *Proceedings of the National Academy of Sciences*, *119*(52), e2207499119.

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65-78.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: developingwhat'andwhere'systems. *Trends in Cognitive Sciences*, *2*(1), 10-18.

Lewis, D. (1972). General semantics. *Semantics of natural language*, 169-218.

Lewis, D. (1986). *On the plurality of worlds* (Vol. 322). Blackwell Oxford.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition*, *96*(1), 67-92.

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics*, *67*, 324-334.

Phillips, J. S., & Kratzer, A. (2022). Decomposing modal thought.

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. MIT press.

Redshaw, J., & Suddendorf, T. (2016). Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology*, *26*(13), 1758-1762.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*(2), 129-134.

Scholl, B., & Leslie, A. (1999). Explaining the infant's object concept. *What is cognitive science*, 26-73.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1-2), 1-46.

Stalnaker, R. C. (1984). Inquiry.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences*, *12*(8), 291-297.

Turan-Küçük, E. N., & Kibbe, M. M. (2024). Three-year-olds' ability to plan for mutually exclusive future possibilities is limited primarily by their representations of possible plans, not possible events. *Cognition*, *244*, 105712.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature neuroscience*, *5*(6), 598-604.