Final Pre-publication Version
Philosophical Transactions of the Royal Society B (2022)
Please see published version at:
https://royalsocietypublishing.org/doi/epdf/10.1098/rstb.2021.0355

A Context-Sensitive and Non-Linguistic Approach to Abstract Concepts

Peter Langland-Hassan and Charles P. Davis

Abstract: Despite the recent upsurge in research on abstract concepts, there remain puzzles at the foundation of their empirical study. These are most evident when we consider what is required to assess a person's abstract conceptual abilities without using language as a prompt or requiring it as a response—as in classic non-verbal categorization tasks, which are standardly considered tests of conceptual understanding. After distinguishing two divergent strands in the most common conception of what it is for a concept to be abstract, we argue that neither reliably captures the kind of abstraction required to successfully categorize in non-verbal tasks. We then present a new conception of concept abstractness—termed Trial Concreteness—that is keyed to individual categorization trials. It has advantages in capturing the context-relativity of the degree of abstraction required for the application of a concept and fittingly correlates with participant success in recent experiments.

1. Introduction

The scientific study of abstract concepts has greatly expanded in recent years (Bolognesi & Steen, 2018; Borghi, Barca, Binkofski, & Tummolini, 2018; Yee, 2019). Yet clarity is still needed on two central fronts. First, there are questions concerning the nature of abstract concepts themselves: what makes them abstract? What is it that makes one concept *more* abstract than another? Second, there are questions concerning the proper means for studying abstract concepts. Through what kinds of experiments can we confirm that someone understands or is using one or another abstract concept?

The second question is especially pressing in settings where researchers are keen to understand the relationship between abstract concepts and language (Borghi et al., 2019; Dove, 2018; Dove, Barca, Tummolini, & Borghi, 2020; Langland-Hassan, Faries, Gatyas, Dietz, & Richardson, 2021), and, relatedly, whether abstract concepts may be possessed by non-human animals (Penn & Povinelli, 2007) or deep learning artificial intelligence algorithms trained exclusively on non-linguistic stimuli (Buckner, 2018). To assess whether language supports, or is even required for, the use of abstract concepts, researchers can benefit from experimental paradigms that do not engage language comprehension or production simply as a matter of their explicit methods. However, the question of whether and how such tests can be created depends in part on the first nest of questions, concerning the nature of abstract concepts themselves. This paper addresses both questions simultaneously as a means for making progress on the second, methodological one.

In the first part of this paper (Section 2), we distinguish two strands in the most common conceptions of abstract concepts that mark quite distinct ideas about what makes a concept abstract. Namely, a concept may be considered abstract to the extent that its referents are very diverse (particularly in their perceptible properties); or, alternatively, a concept may be considered abstract insofar as its referents are simply imperceptible. We then discuss why such conceptions of abstractness—despite their intuitive appeal—may fail to track the kinds of concept features and conceptual abilities most will want measured by ratings of abstractness. We then develop, in Sections 3 and 4, a context-sensitive conception of abstractness called *Trial Concreteness* that is sensitive to the ways in which the level of abstraction required by a conceptual task varies as a feature of its context. There we draw on new analyses of previously published data to reveal an interesting trend: participant success on non-verbal semantic memory tasks does not differ as a function of the abstractness of the concept tested—at least, not when we understand abstractness in either of the two traditional senses. By contrast, *Trial Concreteness* both correlates with participant success on a recent study and captures what we should want from a measure of concept abstractness. We conclude that the study of abstract concepts will benefit from placing greater emphasis on context-relative measures of abstractness (such as Trial Concreteness) over more traditional measures (such as word concreteness) that assign degrees of abstractness to concepts regardless of their context of use. A further advantage of Trial Concreteness—connected to the second question above—is that it is a measure of abstractness that is not inherently tied to words and word ratings, making it especially suitable for investigating the relationship between abstract conceptual abilities and language.

2. *Two strands in the notion of an abstract concept*

What is an abstract concept? Aiming at a general and uncontroversial characterization, Borghi et al. (2017) describe abstract concepts (such as FANTASY, FREEDOM, and JUSTICE) as "lacking bounded and clearly perceivable referents" and as being "more detached from sensorial experience than concrete ones" (p. 263). Borghi et al. (2018) also note that abstract concepts "do not possess a single and perceptually bounded object as referent," and have a content that is "more variable both within and across individuals" (p. 1). While adequate as initial characterizations, these thumbnail definitions are compatible with two very different ways of understanding what it is to be an abstract concept (both are also referenced in Borghi & Binkofski (2014, pp. 3-4)). On one hand, these characterizations contain claims about the *diversity of the referents* (or category members) of abstract concepts. Abstract concepts are said to lack a *single object* as referent, having referents that are not "perceptually bounded" and "more variable." On the other hand, our ability to actually perceive the referents is called into question: the referents of abstract concepts are not "clearly perceivable" and are "more detached from [our] sensorial experience." In a similar vein, Barsalou (2003, p. 1178) characterizes abstract concepts as concepts that are "detached from physical entities" and "more associated with mental events," suggesting, ambiguously, that their referents are either perceptually dissimilar or, perhaps, not even physical entities at all.

As noted by Borghi & Binkofski (2014, p. 3-4), these two strands can easily come apart. Consider the first strand whereby abstractness has to do with the diversity of a concept's referents. Most concepts will be abstract in that sense to some degree, insofar as they apply to things with considerable perceptible differences among them, abstracting away from those differences to unite the members under some similarity. The concept DOG, for example, applies to animals of many different shapes, colors, and characters, so long as they are all dogs. Despite this abstract character to DOG, it is not normally offered as an example of an abstract concept. Why not? One reason is that there are concepts whose referents are far *more* diverse in their qualities—concepts such as MAMMAL, ANIMAL, and

ORGANISM, for instance. As we head upward in this hierarchy—from dog, to mammal, to animal, to organism—toward ever more inclusive categories, their members become "more variable" and lack a "single object" (and, perhaps, are not "perceptually bounded," though we are unsure what that amounts to). Thus, while most concepts are abstract at least in the sense that they abstract away from differences in their referents to unite them under some shared feature, some do so to a very high degree and are in that sense *highly abstracted*. The concept OBJECT is a paradigmatic example of a concept that is highly abstracted in this sense. Thus, in speaking of "abstract concepts," we may have concepts that are highly-abstracted-from-perceived-features of their referents in mind and may contrast them to "more concrete" (i.e. less abstracted) concepts, such as DOG and TABLE (Rosch, 1978; Yee, 2019). Bolognesi et al. (2020) introduce the term *specificity* to refer to the same property, whereby one concept may be more or less inclusive than another, with the least inclusive concepts being the highest in specificity (and requiring the least abstraction). (See also Davis *et al.* (2020), for the related if distinct conception of *situational systematicity*.)

Now consider the second strand in the above definitions of "abstract concept": the idea that the referents of abstract concepts are not clearly perceptible (if perceptible at all). It is clearly wrong to say that the referents of all highly abstracted concepts are not clearly perceivable. Organisms are, in general, easy to perceive. Dogs, chickens, and hippopotami are no *more* perceivable than organisms, for the simple reason that they *are* organisms. Nor is it true to say that we can only perceive an organism "to some extent." To clearly see a dog is to fully perceive an organism. So, if our concept ORGANISM is detached from sensory experience, it is only in the sense that our perceptions of organisms are highly variable. It is *not* in the sense that organisms are not clearly perceivable, or only perceptible "to some extent." In short, the referents of a concept may be highly abstracted from past experience without themselves being difficult to perceive. (See Löhr (2022) for related points.)

By contrast, the referents of other concepts standardly held to be abstract *do* seem difficult to perceive and are perhaps not perceptible at all. Consider, first, mathematical concepts, such as the concept SEVEN. There is no sense to the question of where the number seven is, or how big or small it is, or how fast it is moving; this is because the number seven lacks any spatiotemporal location. Lacking a spatiotemporal location, the number seven is not perceivable at all and thus not "clearly perceivable." Is the concept SEVEN a *highly abstracted* concept, like ORGANISM? It seems not. We cannot have highly variable perceptions of things we simply cannot perceive. Perception aside, does the concept SEVEN include in its extension many individuals with disparate qualities, abstracting-away from those differences to unite them under the concept SEVEN? Again the answer is no. There is just one referent of SEVEN: the number seven itself.[1]

Endorsing this same notion of abstractness-as-imperceptibility—and distinguishing it from abstractness as a feature of the diversity of a concept's referents—Borghi & Binkofski (2014) note:

> Concepts such as ''freedom'' and ''phantasy''…are not abstract because they are on top on a conceptual hierarchy, but because their referent/s are not concrete objects or entities: they are not visible, manipulable or perceivable through any of our senses (p. 3).

So, on Borghi & Binkofski's understanding of abstractness in the imperceptibility sense, abstract concepts have referents that are "not concrete objects or entities," and not "visible, manipulable or perceivable through any of our senses." Clearly, this is a different sense in which a concept may be abstract than the sense in which ORGANISM and MAMMAL are abstract. Moreover, it may be that some concepts of imperceptible entities are less specific and more general than others, in the sense that there is more diversity among their referents (Bolognesi et al., 2020). For instance, there may be some sense

[1] Of course, there are disparate things we may *associate with* seven, which can be perceived: a group of seven apples, or the Arabic numeral '7', for instance. But in perceiving these things, we are not perceiving the number seven itself any more than we perceive cows when we see the word 'cow'.

in which the referents of BELIEF, CIRCUMSTANCE, and AMORALITY are more diverse than those of SEVEN or MONOTHEISM. This highlights the fact that there are two different notions of abstractness in play that can dissociate. A strategy one may adopt in recognition of these distinct notions (and in fact adopted by Borghi & Binkofski (2014)) is to specify that one is interested in only one or the other conception of abstract concept and to focus one's experimental work and theoretical claims on that conception alone. However, as a matter of disciplinary practice, it is more common to blur the lines between the two, perhaps assuming that they retain some important commonality. For instance, Bolognesi et al. (2020) draw a similar distinction between "Abstraction and Concreteness," finding the two notions "theoretically distinct," with 'abstraction' "describing the construction of conceptual categories starting from experiences," [this being akin to our abstractness as diversity of referents] and 'concreteness' as "describing the perceptibility of the referents designated by given concepts" [this being akin to our abstractness as imperceptibility]. They likewise warn that the "polysemous nature of the term *abstraction* in some cases may suggest that the two variables are conflated" (p. 367).

2.1 Troubles with concreteness ratings

Most notably, we find both notions of abstractness at work in common characterizations of the distinction between concrete and abstract *words*, captured in word *concreteness ratings*. In much of the literature on abstract concepts, the concreteness (or imageability (Cortese & Fugett, 2004)) of a word is considered equivalent to the level of concreteness of the concept it expresses, with abstractness increasing proportionately as concreteness decreases. Indeed, the fact that mental processing is, in some contexts, slower for words with lower concreteness ratings (known as the "concreteness effect") has spurred much of the work on the abstract/concrete concept distinction (Begg & Paivio, 1969; Kounios & Holcomb, 1994). However, when we look at the prompt used to generate some of the most influential concreteness ratings—from Brysbaert et al. (2014)—it appears to trade ambiguously on both of the above notions of abstractness ratings. They present participants with the following prompt:

> Some words refer to things or actions in reality, which you can experience directly through one of the five senses. We call these words concrete words. Other words refer to meanings that cannot be experienced directly but which we know because the meanings can be defined by other words. These are abstract words. Still other words fall in-between the two extremes, because we can experience them to some extent and in addition we rely on language to understand them. We want you to indicate how concrete the meaning of each word is for you by using a 5-point rating scale going from abstract to concrete (p. 906).

At first glance, the prompt seems to cohere mainly with the conception of abstractness as imperceptibility. It explains that concrete words refer to things or actions one can "experience directly through one of the five senses," while the referents of abstract words "cannot be experienced directly." It seems we should expect concreteness ratings to line up with the notion of abstractness as imperceptibility and not abstractness as diversity of category members. However, the prompt adds that some words "fall in-between" because "we can experience them to some extent." This is a crucial addition. Without it, concreteness ratings should be binary in nature—1 for if the referent is perceptible with the five senses and 0 if not. In reality, they fall on a spectrum of 1 to 7 (with 7 as most concrete), in keeping with the idea that perceptibility can somehow come in degrees. At a conceptual level, as soon it is suggested by the prompt that the concreteness of a word can fall on spectrum, the measure is asking participants to rate words on some feature *other than* whether the referents for words are perceptible. Thus, those interested in abstract concepts in the imperceptibility sense should hesitate to adopt concreteness ratings as measures for the kind of concept abstractness that interests them. (See Löhr (2022) for related concerns about concreteness ratings.)

However, there is *also* reason for those interested in abstractness in the diversity-of-referents (or "concept specificity") sense to be unsatisfied with concreteness ratings as a proxies for levels of abstractness. In a recent analysis, Bolognesi et al. (2020) compared word concreteness ratings to their proprietary measure of word specificity, where specificity was calculated based on the place a word fell within hierarchies of more and less inclusive nouns in the WordNet database (an electronic database that charts a variety of semantic relations among words (Fellbaum, 1998)). We can expect a word's specificity to mesh fairly well with abstractness in the "diversity of referents" sense, for the simple reason that any word that is less specific than another (e.g. 'furniture' is less specific than 'chair') will have the referents that the other has, plus some. However, Bolognesi et al. found only small-to-moderate correlations between concreteness ratings and specificity: r=0.263, r=0.267, and r=0.354, for each of three slightly different ways of calculating specificity ratings from WordNet placements (p. 372).

In short, because distinct notions of abstractness are invoked in the prompts used to generate concreteness ratings, the ratings themselves fail to track closely either sort of abstractness. This does not entail that concreteness ratings do not latch on to some interesting psychological distinction or other. There are, after all, the well-established "concreteness effects." The point is simply that word concreteness does not line up well with either of the senses of concept abstractness that we have distinguished. Thus, to the extent we wish to take those notions of abstractness as topics of scientific study, we should be wary of using concreteness ratings as measures for them.

We will defend, below, an alternative measure of abstractness—one that we think provides a better measure of the cognitive capacity relevant to concept abstractness in the diversity of referents sense. We close this section with some suggestions for why it will be more profitable to focus research on abstractness in this "diversity of referents" sense, as opposed to abstractness in the imperceptibility sense.

2.2 Troubles with abstractness as imperceptibility

Earlier we noted that perceptibility is, strictly speaking, all-or-nothing. There is no literal sense to the claim that a certain type of thing is only perceptible to a certain degree. We may say of a certain thing—a horse off in the distance, clouded in fog—that it is only somewhat perceptible, or barely perceptible. But this does not make horses in general any less perceptible than, say, tables. For any type of thing at all, there will be a fact as to whether in normal (unobscured) circumstances it is the sort of thing we can perceive with one of the five senses, or not.[2] Nevertheless, despite the fact that there is such an answer, in very many cases we simply won't know what the answer is! To see this, we can begin with clear cases on each side. There are things it is clear we cannot perceive with the five senses, such as numbers, quarks, and the electromagnetic force. And there are many clear cases of perceptible entities, including dogs, chairs, trees, and trains. But there are deceptively many things in between. Are conferences things we can perceive? Are universities? If we can't perceive them, how do we navigate our way to them? If we can perceive them, what shape and color do they have? What about generosity and senators? If we are not literally seeing generosity when we view a generous act, are we, likewise, not literally seeing a senator when we see a person who is a senator? Similar questions can be raised for a vast number of concepts that we might like to investigate. Taking such questions seriously may lead us to rethink what seemed to be the clear cases. Supposing that we cannot literally see a conference—but, instead, only see various things associated with them—in what ways are dogs and

[2] Perhaps an exception are things that are on the border between being too small and just large enough to perceive. A different issue is the degree to which things differ in *perceptual strength*, understood as the degree to which the referents of a concept tend to be experienced multimodally (e.g., Lynott et al., 2020). Whereas there are no doubt gradations in perceptual strength among things which are perceptible, perceptibility itself is, by-and-large, binary.

chairs importantly different? If something that is not a dog can have all the superficial characteristics associated with dogs (just as something that is not a conference might have all the superficial characteristics of something that is not one) does this entail that we do not literally perceive dogs either, but simply infer their presence based on perceptible properties with which they are associated? Giving principled answers to such questions requires a well-developed philosophical theory of perception—one that will itself be controversial.

In view of these difficulties in delineating the realm of the (literally) perceptible, we advise focusing research on the better-grounded notion of abstractness as-diversity-of-referents. We think this notion of abstractness also provides much of what is wanted from the abstractness-as-imperceptibility notion. Much of the interest in the latter stems from empiricist views that see concept-acquisition as arising out of perceptual contact with the referents of a concept—where concepts are seen as "grounded in" those past acts of perception. It is then a serious question for such views how we come to have concepts of things that are imperceptible. Such questions are likewise pressing for any concept whose referents are extremely diverse in their perceptible features. After all, if it is perceptual contact with the things that is supposed to explain how we acquire concepts of them, and if the perceptible characteristics of the things are extremely diverse and unsystematic, then it is not clear that perception is any better placed to explain our concepts of those things than it is to explain concepts of strictly imperceptible items. From the perspective of theories seeking to ground concepts in perception, a maximally inclusive concept such as OBJECT poses an equal if not greater challenge than a concept of a strictly imperceptible entity, such as ELECTRON or MEANING. In both cases, we are interested in how concepts of a certain kind of thing are acquired through having perceptual experiences that are not themselves similar in any obvious way. In short, we are interested in the ability to abstract away from—to in some sense *ignore*—salient differences between things in order to cognitively grasp some genuine commonality. Having identified this as the central question of interest with respect to abstract concepts, the balance of this paper makes the case for shifting to a context-sensitive assessment of the ability for abstraction—one which keys the level of abstraction shown by a participant not to the concept exploited, but to the concept together with the context.

3. *Categorization, semantic memory, and non-verbal uses of abstract concepts*

Last section we distinguished two distinct conceptions of abstractness related to concepts: abstractness understood as diversity in the perceptible features of the concept's referents, and abstractness understood as imperceptibility of the referents. We noted how the two are run together in measures of word concreteness and then recommended focusing on abstractness as diversity among referents, in part due to difficulties in clearly delineating the perceptible from the imperceptible. We now want to return to the two distinct conceptions of abstractness in order to focus on a problem in assessing the relation between them and a person's language abilities. Mixing and matching the two notions, we can generate four classes of abstract concept[3]:

a) *Minimally abstracted concepts of perceptibles*. These are concepts whose referents are perceptible and bear relatively many perceptible similarities. They include paradigmatically concrete concepts such as COW, TABLE, CUP, and ROSE.

[3] See Bolognesi *et al.* (2020) for a similar four-way grouping of concepts. In their framework, abstractness as diversity among referents is equated with a word's "specificity," and abstractness as imperceptibility is equated (problemantically, we think) with a word's concreteness ratings. See also Hoffman et al., 2013, who, in place of the "specificity" notion, emphasize instead the semantic diversity of the linguistic *contexts* in which words appear.

b) *Highly abstracted concepts of perceptibles.* These are concepts whose referents are perceptible and bear relatively few similarities. They include concepts such as OBJECT, ORGANISM, PLANT, and ARTIFACT.
c) *Minimally abstracted concepts of imperceptibles.* These are concepts whose referents are imperceptible and bear relatively many similarities (sometimes because the referent is a single individual). They include concepts such as SEVEN, ELECTRON, and PYTHAGOREAN THEOREM.
d) *Highly abstracted concepts of imperceptibles.* These are concepts whose referents are imperceptible and bear relatively few similarities. They include concepts such as VALUE, THEORETICAL ENTITY, and MENTAL STATE.

We should not expect bright lines between the classes, as at least one of the factors on which they are based—the similarity among the referents of the concept—is clearly a matter of degree. Instead, these four classes mark regions in a continuous space. Second, while we assume that being perceptible is an all-or-nothing affair—and thus does not come in degrees—we also noted that it is not always clear which things are in fact perceptible and which are merely inferred on the basis of, or associated with, certain perceptions.

Nevertheless, this four-way distinction in kinds of concepts allows us to formulate the key question we want to explore in the balance of this paper. If there are important differences in how categorization occurs for abstract as opposed to concrete concepts—and in the kind of cognitive challenges such categorizations present—we should expect to see those differences when comparing performance on class (a) concepts (minimally abstracted concepts of perceptibles) and class (d) concepts (highly abstracted concepts of imperceptibles). Yet, drawing on new analyses of previously published data (Langland-Hassan et al., 2021)—presented below—we argue that no such differences emerge. We then offer, as remedy, a proprietary context-relative conception of abstractness—distinct from the two already presented—that both correlates with participant performance and captures what we should want from a measure of abstract concept deployment.

3.1 Testing a variety of abstract concepts non-verbally

A widely endorsed conception of concepts holds that they are cognitive mechanisms that enable categorizations of a certain sort (Prinz, 2002). The ability to group squirrels with squirrels, and batteries with batteries, for example, is assumed to rely upon use of the concepts SQUIRREL and BATTERY, respectively. We can therefore test a person's conceptual abilities by assessing their ability to categorize appropriately. Categorization tasks come in several varieties. We will focus on one very common form here—a pictorial semantic memory task—versions of which we have used in previous studies (Langland-Hassan et al., 2021; Langland-Hassan, Gauker, Richardson, Dietz, & Faries, 2017). Semantic memory is standardly understood as long-term memory for non-personal facts—such as that whales are mammals, or that brooms are for sweeping—where this memory store is one and the same with one's conceptual knowledge (Yee, Jones, & McRae, 2018). The semantic memory task we have used mirrors in its structure the well-known Cactus to Camels test (Adlam, Patterson, Bozeat, & Hodges, 2010; Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000), which is itself the most commonly used *non-verbal* test of semantic conceptual abilities. It requires a participant to match a target image with the one of four other images that is its appropriate match. Selecting the appropriate image is considered evidence of one's facility with the concept that links the target and match.

In one example of a stimulus from the Cactus to Camels test, a picture of a passenger train is the target image, and the four choice images below it are of a motorcycle, a tractor, a semi-truck, and a bus. Correctly matching the train with the bus requires uniting them under the single *category* of mass-transit vehicles. We can, for that reason, call this a categorical semantic memory trial. Note, in this case, that the target and match images depict instances of a referent of the tested concept. This is

possible because mass-transit vehicles are imageable and likewise perceptible, allowing our concepts of them to fall into either class (a) or (b) above, depending how abstracted one takes the concept mass-transit vehicle to be. Also notice the importance of the distractor items to determining which concept is tested. Had the three distractor items not also been vehicles, it could have been the more general concept VEHICLE—and not the more specific concept, MASS-TRANSIT-VEHICLE—that linked the target and match. This tendency of the distractor items to modulate the nature of the concept tested, and the difficulty of the task itself, will be highlighted below.

Importantly, many of the trials in the Cactus to Camels test are not categorical trials, in that they do not ask participants to match depicted instances of a certain kind of thing with each other. Instead, they require one to determine which two items are *associated* with each other—where the associative link can take a variety of forms. In simple versions of this kind of trial, a piano may be matched with a piano bench, or, indeed, a cactus with a camel. In these trials, the target and match have a *thematic* link, consisting in the fact that the two items are commonly found together in a certain kind of setting. The link is thematic and not categorical because (unlike the two mass-transit vehicles) the piano and piano bench, and cactus and camel, are not the same kind of thing and thus, it seems, not members of a single category. Instead, they are simply found together.

Yet it is worth scrutinizing the thematic/categorical distinction (Kalénine et al., 2009; Markman, 1981; Mirman & Graziano, 2012) a bit more carefully. There is, after all, a way to see the piano and piano bench as being the same kind of thing—even members of the same category—if we allow for categories such as *piano-related*. Likewise for the cactus and camel, if we allow for the category *desert-related*. If to be a desert-related entity is to fall into the category desert-related, then the cactus and camel are members of a single category after all. We can further suppose that there are compound PIANO-RELATED and DESERT-RELATED *concepts* corresponding to such categories, with the thematic trials testing those concepts. (Barsalou (1983) argues for a similar notion of *ad hoc* concepts.) After all, if concepts are simply those cognitive mechanisms that enable relevant categorizations, and a person is able to group things as piano or desert-related, then we can justifiably attribute them the concepts PIANO-RELATED and DESERT-RELATED.

An important feature of thematic semantic memory trials, so understood, is that they can test one's understanding of concepts whose referents are (arguably) imperceptible. For to test such a concept, we do not need to picture the (imperceptible) thing to which the pictured items are related. Instead, we need only picture the related items themselves. To see how this can play out, consider another trial from the Cactus to Camels test, where a power plug is shown as the target image, with a candle, flashlight, lantern, and lightbulb as the choice images. Properly matching the power plug with the lightbulb requires use of a concept such as ELECTRICITY-RELATED. To have this compound concept, one must have the concept ELECTRICITY. And so, while electricity cannot itself be pictured—or, in any case, is *not* here pictured—facility with that concept can be tested pictorially.

In the above-mentioned study, we exploited this feature of thematic trials to test a variety of concepts of imperceptible items, matching a microscope and test tube under the concept SCIENCE-RELATED, scales and gavel under the concept JUSTICE-RELATED, and church and menorah under RELIGION-RELATED (to give just a few examples). However, we also included many categorical trials where the pictured items were instances of a traditional category: eyeglasses and a wine glass were linked under GLASS; a Rubik's cube and die were linked under CUBE; a top hat and baseball cap were linked under HAT. (See Supplementary Materials, Appendix A, for pictures of the full set of these trials, used in Langland-Hassan *et al.* (2021).)

Having noted that concepts of both imperceptibles and perceptibles can be tested through the use of pictorial semantic memory tasks, let us return to the question of whether the explanatory challenge presented by class (a) concepts (minimally abstracted concepts of perceptibles) is of a fundamentally different kind than that presented by class (d) concepts of imperceptibles (highly

abstracted concepts of imperceptibles). A natural assumption is that, if there is an important difference in these classes of concepts—either in how they are acquired or in the kinds of representations or cognitive mechanisms underlying their use—that difference should show itself in participant performance. This assumption has risks, of course, as it may be that participants perform equally on the two kinds of task, despite their requiring the use of quite different kinds of concepts and cognitive abilities. Or it could be that, despite their posing different cognitive challenges, all else equal, other kinds of differences in the trials (such as the familiarity of the items) counteracts those differences in ways that obscure them. Nevertheless, having flagged these risks, it remains natural to expect that categorizing in accordance with more abstract concepts presents a greater cognitive challenge than grouping items under less abstract concepts. After all, abstract concepts, with their oft-supposed link to natural language (Borghi et al., 2019; Dove, 2018), are commonly thought to be the exclusive province of human minds. To the extent that we may indeed share less abstract concepts with non-human animals, we should find such concepts easier to master and apply.

Thus, with the above caveats, we should, at a minimum, expect to see performance differences between trials testing concepts of kind (d) and those testing concepts of kind (a). For example, returning to the examples above, we should expect that trials testing the concepts JUSTICE-RELATED, SCIENCE-RELATED, and RELIGION-RELATED will pose a greater cognitive challenge than those testing GLASS, HAT, and CUBE. However, across the 80 trials of the above-cited experiment (Langland-Hassan, et al., 2021), we did not observe any such trend (the full statistical analysis, reported below, was not previously published). For example, in our sample of 146 neurotypical participants who completed the 80 trials online, 93% correctly matched the scales and gavel on the JUSTICE-RELATED trial; 93% correctly matched the church and menorah images on the RELIGION-RELATED trial; and 97% correctly matched the test tube and microscope on the SCIENCE-RELATED trial. By comparison, 83% answered the CUBE trial correctly, 96% answered the HAT trial correctly, and 93% answered the GLASS trial correctly. Moreover, the mean correctness across all 80 trials was only 83% (Langland-Hassan et al., 2021). (See Supplementary Materials, Appendix B, for item-by-item accuracy and VIS/COM data.) Looking just as these three examples, it appears that categorizing in accordance with the most abstract of concepts—*viz.,* highly abstracted concepts of imperceptibles—did not pose a greater challenge than categorizing with respect to the most concrete.

In a fuller investigation of that trend, we return here to the data from Langland-Hassan et al. (2021) to investigate this apparent lack of relationship between the abstractness of a concept and the difficulty of a trial. (This question was not at the forefront of the earlier investigation, so was not then explored.) Carrying out such a calculation requires that each trial has an associated degree of abstractness. In the original study, we assigned such a rating using the following method. (Full details on the procedure are included in Langland-Hassan et al. (2021)). After participants (n participants per trial = 146) in the norming study selected the item they thought best goes with the target item, they were asked to write the one word that best captures what the two have in common. The assumption was that this word would reliably express the concept used in their categorization. We then determined which such word was given most often in cases where the trial was answered correctly. This became the *linking word* for the trial. Thus, for example, 'science' was the linking word for the trial that matched a test tube with a microscope, because 'science' was the word most often given as the word that best captures what the target and match have in common. We then assigned a concreteness/abstractness rating to each trial corresponding to the Brysbaert et al. (2014) word concreteness ratings, discussed above. We referred to this measure as 'Concept Concreteness,' with trials low in Concept Concreteness

(and thus low Brysbaert et al. (2014) concreteness ratings) being rated more abstract than those with high concreteness ratings.[4]

We have discussed above the ways in which the Brysbaert et al. concreteness ratings gesture, somewhat hazily, at two different conceptions of abstractness. As we are now considering whether tests of concepts that are abstract in *both* of those senses—i.e., concepts of class (d)—are more challenging for participants than those that test concepts that are abstract in neither sense, the ratings remain relevant to our purposes (even if they remain imperfect measures of each construct). Assigning a concreteness/abstractness rating to each trial—via the concreteness rating of its Linking Word—allowed us now to assess interactions between abstractness level (so conceived) and the difficulty of the trial. Across the 146 participants and 80 trials, we found no reliable correlation between participant accuracy and the abstractness (understood as word-concreteness) of the associated concept (r = .088). To understand why there would be no such correlation—and what this has to do with abstractness generally—it is useful to consider it in contrast to a separate measure that *did* strongly predict a participant's ability to select the correct answer. This is a proprietary, context-sensitive measure we called *Trial Concreteness* (to be distinguished from the above-mentioned, and Brysbaert-concreteness-norms-related, *Concept* Concreteness). We now turn to explaining this measure and its rationale.

4. Trial Concreteness: Toward a context-relative conception of conceptual abstractness

The measure of Trial Concreteness was motivated by the twin insights that 1) what is interesting about abstract concepts is that their use requires abstracting away from diversity in the concept's referents and that 2) that the amount of abstraction that a specific categorization trial requires for its solution depends critically on the nature of the distractor items (and is in that way context-relative). To make the latter explicit, consider Figure 1a, a purely hypothetical trial in the format of the pictorial semantic memory tasks used in Langland-Hassan et al. (2021): at the center top is a target image of a German Shepherd. Below it are four choice images, with the Poodle being the intended match (the target and match being linked under the concept DOG). At first blush, it might seem that the difficulty and abstractness of the trial will mainly have to do with the abstractness of the concept DOG. As 'dog' is a paradigmatically concrete word, it may be unsurprising that the trial appears relatively easy.

[4] In the original study, we also considered ratings of imageability (Cortese & Fugett, 2004) and Sensory Experience Ratings (Juhasz & Yap, 2013) as alternative measures of Concept Concreteness. However, the Brysbaert et al. (2014) ratings are more comprehensive and more commonly used; so we focus on them here.

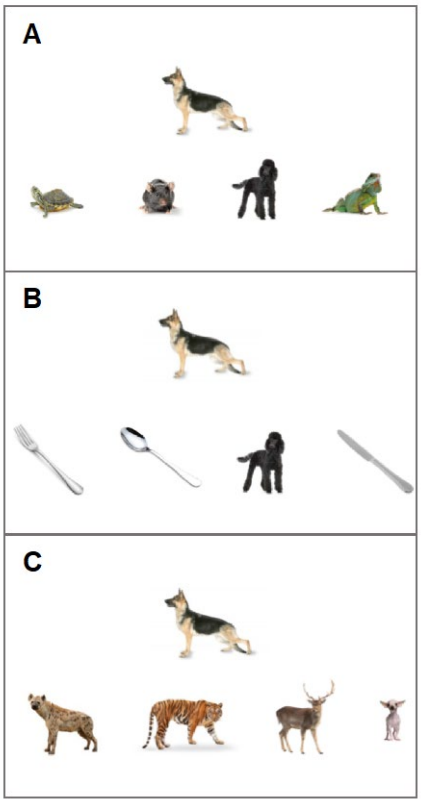

Figure 1. A set of hypothetical trials illustrating the role of context in facilitating use of the same "linking" concept DOG. *In (A) the correct answer, poodle, is flanked by animal distractors, albeit ones which are visually dissimilar and do not occur in similar contexts as dogs; in (B) the correct answer, poodle, is flanked by culinary utensils which are even more visually dissimilar from the German Shepherd; and in (C) the correct answer, chihuahua, is flanked by distractors which are far more visually similar to the German Shepherd, thus requiring a greater degree of abstraction to elicit a correct response.*

Yet now consider two alternative hypothetical trials (Figures 1b and 1c) with the same target image of a German Shepherd and where the uniting concept is still DOG, yet where changes to the choice images substantially alter the nature of the task. In Figure 1b, the distractors are all culinary utensils and the Poodle remains as the match. In Figure 1c, the distractors are all other animals—some, such as the hyena, are quite visually similar to the German Shepherd—and the match is a Chihuahua. It is easy to see that these three trials will not be equally difficult and, intuitively, do not require the same degree of abstraction. Matching the German Shepherd with the Poodle is far easier when the distractors are utensils than when they are other animals; and matching the German Shepherd with the Chihuahua—as dogs—is much harder when the distractor animals look more like the German Shepherd than does the Chihuahua. This increasing difficulty corresponds to how many of the visual similarities there are between the target and match that must be ignored, and how many visual differences there are between the target and match that must be abstracted-away from. As the number of each increases, the trial becomes more abstract.

4.1. Calculating context-relative Visual Similarity (VIS)

To formalize this trial- and (distractor-) relative feature of abstractness, we included a norming stage where a separate set of participants (n = 191) were asked to sequentially rate how visually similar the target is to each of the four choice images. Using Figure 1a as a (hypothetical) example, a

participant would be asked to give a rating from 1 to 7 (with 7 corresponding to "they look exactly alike" and 1 to "they look nothing alike") to the pair of the German Shepherd and turtle; then, they would similarly rate the pair of German Shepherd and mouse, and so on, for the remaining two pairs of target and choice image. A Visual Similarity rating for the trial could then be calculated by subtracting from the mean score of the target and match the score of the next highest rated pair. So, if the mean visual similarity score of the German Shepherd and Poodle pair, in Figure 1b, was 6, and the next highest visual similarity score of another pair (e.g., the German Shepherd and mouse) was 4, this would result in a Visual Similarity rating of 2 for the trial. The lower this score, the more difficult it should be to correctly answer the trial by noticing perceptible similarities between the target and match. Accordingly, we would expect the Visual Similarity rating for the trial in 1b to be significantly higher than that of 1a, and, likewise, the Visual Similarity of 1c to be lower than that of 1a. In cases where the target and match were judged to be less visually similar than the target and one of the other choices, the trial would get a negative Visual Similarity score.

The hypothetical examples of Figures 1a-c show how such a rating will vary depending on the choice images, even in cases where the target and match are linked under the *same concept* each time—a concept with a *single* related word concreteness rating and with a single specificity score (in Bolognesi et al.'s sense, discussed above). Importantly, as we will see below, this Visual Similarity rating did significantly correlate with participant success at selecting the matching image (see Figure 2). This bolsters the intuitive sense that Visual Similarity is better tracking the degree of abstract thought required for the categorization than related word concreteness ratings.

4.2 Calculating context-relative Common Setting Association (COM)

Recall first that word concreteness ratings, in the way they are solicited, run together the two notions of abstractness we have distinguished: abstractness as imperceptibility, and abstractness as diversity among instances. We have just seen one reason to think that the kind of abstraction required by categorization tasks—which, again, are tasks that show evidence of *concept* possession—is not well tracked by notions of abstractness that are fixed to individual words or concepts (as opposed to the context of individual *trials*). A second reason for skepticism about the utility of word-related abstractness ratings in the context of categorization tasks is the considerable ease with which participants answered many trials with words deemed relatively non-concrete (such as the SCIENCE-RELATED, RELIGION-RELATED, and JUSTICE-RELATED trials mentioned above). Why were these trials, which query concepts whose referents are *not* perceptible, and where the target and match were not particularly visually similar (relative to the distractors), answered correctly at comparatively high rates? The reason is that the visual similarity between target and match is not the only factor that modulates ease and degree-of-abstraction required to correctly categorize. Trials are also less demanding when the target and match are strongly associated together, due to their being commonly found together. Importantly, this is so *whether or not* the concept linking the items is a concept of something perceptible.

Here, the SCIENCE-RELATED trial provides a good example. In this trial (Fig. 2), a microscope was the target image and a test tube its match. The four choice images were a glass vase, a wine glass, and a glass bottle. This trial received only a 0.65 Visual Similarity rating, reflecting the fact that it would not be easy to solve simply by latching on to relative visual similarities between the target and match. However, microscopes and test tubes are very closely associated, relative to the microscope and other items. This is because microscopes and test tubes are frequently found together in a common (science-related) setting. The fact that they are so commonly found together means that the trial requires less abstraction than one where the target and match are neither visually similar nor commonly found together. In what sense does it involve less abstraction? In this case, the sort of abstraction required is not abstraction away from the perceptible features of the target and match, but away from regularities

in past perceptual episodes. In the case of the trial in Figure 2a, there are very many situations where the test tube and microscope are found together in a common setting, compared to the microscope and distractors. Thus, to answer the trial correctly, participants need not abstract away from those common pairings to grasp some other shared feature of the target and match. They can instead be guided by a relatively automatic association between the two items—one plausibly grounded in past acts of perception where they appeared together.

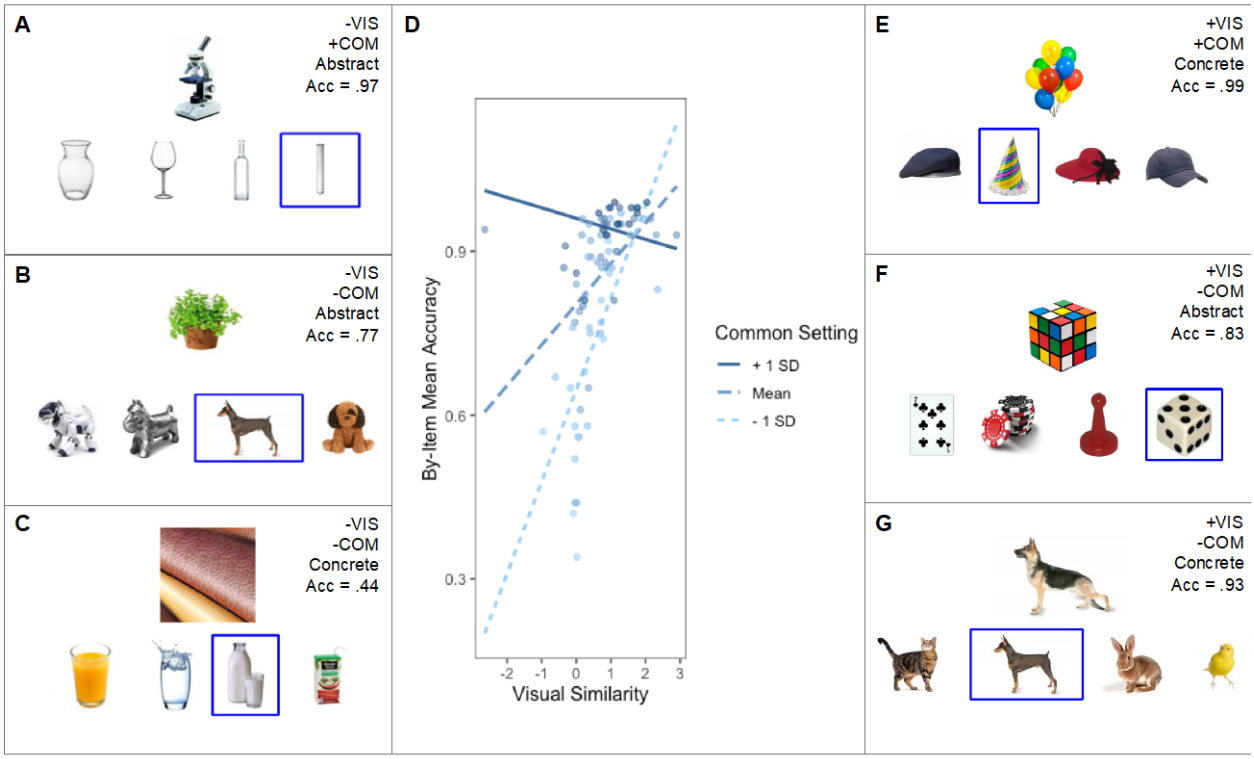

Figure 2. Example trials varying in Visual Similarity (VIS), Common Setting (COM), the degree to which the "linking" concept is canonically concrete or abstract, according to the Brysbaert et al. (2014) norms, and average participant accuracy on that trial (Acc). (A) A depiction of the SCIENCE trial, which is a canonically abstract concept low in VIS but high in COM; (B) the ALIVE trial, also canonically abstract but low in both VIS and COM; (C) the COW trial, which is canonically concrete but low in both VIS and COM; (E) the PARTY trial, which is canonically concrete and high in both VIS and COM; (F) the CUBE trial, which is abstract yet high in VIS and low in COM; and (G) the DOG trial, which is highly concrete, high in VIS, and low in COM. In (D) we depict the interaction between VIS and COM in predicting participant accuracy, with the effect of VIS on accuracy increasing in strength as COM decreases; SD = standard deviation.

As with Visual Similarity, we assigned a rating to each trial reflecting the (choice-relative) degree to which a trial could be solved by latching on to past regularities in whether the target and choice were perceived together in a common setting. Echoing the norming procedure by which Visual Similarity Ratings were acquired, we asked participants (n = 197) to sequentially rate how frequently the target and each of the choice items are found together in a common setting—with 7 being "always found together" and 1 being "never found together." An overall Common Setting score was then calculated for the trial by taking the mean score for the target and match and subtracting from it the mean score for the next highest rated pair of target and choice image. For the SCIENCE-RELATED trial (Figure 2a), this resulted in a relatively high Common Setting score of 4.24.

4.3 Trial Concreteness as a general, context-relative measure of abstractness

To demonstrate that Visual Similarity and Common Setting—two context-sensitive measures of abstractness—better predict participant performance than canonical concreteness ratings, we performed a two-step linear regression model predicting participant accuracy. In the first step, Brysbaert et al. (2014) concreteness ratings as well as our Visual Similarity and Common Setting metrics were entered as predictors. Based on the above discussion, we anticipated that concreteness ratings would not reliably predict trial accuracy, whereas Visual Similarity and Common Setting would each have reliable and independent effects on participant behavior. This is precisely what we found. There was no effect of concreteness rating on accuracy (est = .004, SE = .014, t = 0.30, p = .76), but each unit increase in Visual Similarity was associated with an accuracy increase of .07 (SE = .018, t = 4.01, p < .001) and each unit increase in Common Setting was associated with an accuracy increase of .05 (SE = .010, t = 5.09, p < .001).

Because Visual Similarity and Common Setting scores can vary independently—both tracking an ability to abstract away from perceptually salient features—in Langland-Hassan *et al.* (2021) we summed the two into a single score—called *Trial Concreteness*—intended to express the general abstractness of the trial. As a validation of this combined measure of abstractness, we here added, in step 2 of the model, the interaction between Visual Similarity and Common Setting to the step 1 model with only the individual predictors, and demonstrate an interactive effect on trial accuracy (est = -.06, SE = .013, -4.97, p < .001), whereby the effect of Visual Similarity is most pronounced when Common Setting is low (see Figure 1d). By way of a model comparison between the step 1 main effects model and the step 2 interaction model, we demonstrate that Trial Concreteness was a better predictor of trial accuracy than the main effects alone ($F(1, 75) = 24.70$, $p < .001$).

In an example of a low Trial Concreteness trial, a potted plant was the target image and the match a Doberman Pincer (Figure 2b). The three distractor images were of a robot dog, a metal dog figurine, and a dog-like stuffed animal. In norming, the target and match were paired under the linking word 'alive'. Because the target and match were not especially visually similar, relative to the other choices, nor found together commonly, relative to the other choices, the trial had a relatively low Trial Concreteness rating (0.72). Participants found this trial challenging, answering correctly on 77% of trials. Of course, the concept ALIVE is intuitively abstract and indeed seems abstract in both the abstractness-as-imperceptibility and abstractness-as-diversity-among-instances senses. (It has a Brysbaert et al. (2014) concreteness rating of 3.14, where 5 is the highest possible score.) Thus, it may seem unsurprising that this trial would likewise result in a low Trial Concreteness rating. Yet we have already seen that other trials involving paradigmatically abstract concepts—such as SCIENCE, RELIGION, and JUSTICE—did *not* have low Trial Concreteness ratings. And neither is it necessary that a trial tests a traditionally abstract concept in order for it to have a low Trial Concreteness rating. Consider the trial displayed in Figure 2c, where a piece of leather, as target item, is matched with a bottle of milk—both being cow-related (with 'cow' being the normed linking word). The distractor images are of other beverages: orange juice, water, apple juice. Even though 'cow' is a paradigmatically concrete word (Brysbaert et al. (2014) concreteness rating = 4.96), this trial had a very low Trial Concreteness rating (0.00), suggesting that a high degree of abstraction both from perceptual differences, and from regularities in past perceptions, is required to answer it correctly. And, as one might expect, this trial posed a significant challenge to participants, being answered correctly at a rate of only 44% (where chance is 25%). Additional examples of how different degrees of Trial Concreteness predict participant accuracy are depicted in Figure 2e-g.

Thus, we see a double-dissociation between trials that require categorization in accord with paradigmatically abstract concepts, on the one hand, and trials with low Trial Concreteness, on the other. A categorization task can test a concept—such as SCIENCE—whose word has a low degree of concreteness (Brysbaert et al. rating = 2.79), while having a high Trial Concreteness rating (4.24) (as in

the SCIENCE-RELATED trial, Figure 2). And a trial can test a concept—such as COW—whose word has a high degree of concreteness, while having a low Trial Concreteness rating (0.0) (as in the COW-RELATED trial, Figure 4).

Finally, it is clear that the answer to some trials will be more obvious than others and that, for some trials, the correct answer may be ambiguous. One might ask how we can know that the predictiveness of Trial Concreteness is related specifically to the level of abstraction required by a trial, as opposed to the sheer ambiguity of the trial. In response, note first that there are two kinds of potential ambiguity: a trial may be ambiguous simply because the answer is not immediately obvious; or, it may be ambiguous because there is more than one plausible answer. The latter sort of ambiguity was ruled out through a norming process. Any trial where an answer other than the intended correct answer was given by more than 25% of participants was excluded from the final set of trials used in the above analyses (Langland-Hassan et al., 2021). With respect to the former form of ambiguity—which is simply a matter of a trial's not having any immediately obvious answer—we should expect low levels of Trial Concreteness to result in trials where the answer is less salient. This lack of salience is explained by the fact trials low in Trial Concreteness require a greater degree of abstraction away from the salient perceptual features of the stimuli and from past regularities in perception (whereby some items are typically found together). It remains possible that some factor other than the amount of abstraction required to answer correctly is tracked by Trial Concreteness and that this is what accounts for the relative difficulty of some trials. However, given the particular way in which Trial Concreteness is calculated, it is not clear what this other difficulty-increasing factor might be or why it would be tracked by Trial Concreteness.

5. Implications and conclusions

We have seen reason to prefer ratings of Trial Concreteness over more traditional ratings of concept abstractness when using categorization tasks to test abstract conceptual abilities. It is important to note that Trial Concreteness, as we have defined it, has no direct relation to the perceptibility of the referents of the concept that links the target and match. The tested concept may be of something paradigmatically perceptible (like cows), or not (like science and justice), while the trial remains low—or high—in Trial Concreteness. Thus, advocating Trial Concreteness as a more meaningful measure of what we are trying to assess when testing abstract concept use means that we have substantively shifted focus away from what many have studied when aiming to study abstract concept use. Trial Concreteness is purposely out of step with the question of the perceptibility of the concept's referent.

What about the other traditional conception of concept abstractness, as diversity among instances of the referent? Here, too, there is a significant tension with Trial Concreteness. When calculating a VIS score, what matters is not the visual similarity among *all possible* instances of the concept's referents; it is, instead, the visual similarity of *two* members, relative to the visual similarity of the target and distractors. Further, as revealed by the contribution of COM scores to Trial Concreteness, there are cases where perceptible similarities among instances of the referent are not relevant to the abstractness of the trial, because, despite the target and match's being perceptually dissimilar (even relative to the other choices), their close association within a context entails that little abstraction away from regularities in past perception is needed to perform the correct categorization.

Another way to express this theoretical conclusion is that, in practice, there is little value in speaking of the abstraction required by use of a particular concept. That is, we see minimal value in assigning a fixed degree of abstractness to any individual concept, given that this value will itself be a kind of averaging across very many contexts where varying degrees of abstraction were required for the behaviors we associate with use of the concept. It remains true that some concepts will, on average,

require more abstraction for their use than others. In that sense, which averages across all usages, we can say that one concept is more abstract (in tending to require, for its use, more abstraction away from differences) than another. Concepts of more inclusive categories (such as MAMMAL) will *tend to* require more abstraction than narrower categories (such as DOG). Nevertheless, depending on the context, these differences may be reversed on an any particular task that aims to test facility with a concept. If we wish to investigate the support language may provide for abstract conceptual abilities, it is more useful to rate the degree of abstraction required by a particular use of a concept in context. We can then assess how one's linguistic capacities interact with one's abilities to deploy concepts in contexts that require a high level of abstraction.

Whether there are other categorization tasks—or other tests of concept possession generally—that favor more traditional conceptions of abstractness (and their related ratings, such as the Brysbaert *et al.* (2014) concreteness norms) of course remains an open question. We hope to have added some complexity to those questions by clarifying the two distinct strands in the received view of concept abstractness and showing that a context-sensitive, trial-relative measure of abstractness may often tell us more of what we want to know.

References

Adlam, A. L., Patterson, K., Bozeat, S., & Hodges, J. R. (2010). The Cambridge Semantic Memory Test Battery: detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase, 16*(3), 193-207. doi:10.1080/13554790903405693

Barsalou, Lawrence W. (1983). Ad hoc categories. *Mem Cognit, 11*(3), 211-227. doi:10.3758/BF03196968

Barsalou, Lawrence W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1435), 1177-1187. doi:doi:10.1098/rstb.2003.1319

Begg, Ian, & Paivio, Allan. (1969). Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior, 8*(6), 821-827. doi:https://doi.org/10.1016/S0022-5371(69)80049-6

Bolognesi, Marianna, Burgers, Christian, & Caselli, Tommaso. (2020). On abstraction: decoupling conceptual concreteness and categorical specificity. *Cognitive Processing, 21*(3), 365-381. doi:10.1007/s10339-020-00965-9

Bolognesi, Marianna, & Steen, Gerard. (2018). Editors' Introduction: Abstract Concepts: Structure, Processing, and Modeling. *Topics in cognitive science, 10*(3), 490-500. doi:10.1111/tops.12354

Borghi, Anna M, Barca, Laura, Binkofski, Ferdinand, Castelfranchi, Cristiano, Pezzulo, Giovanni, & Tummolini, Luca. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of life reviews, 29*, 120-153.

Borghi, Anna M, Barca, Laura, Binkofski, Ferdinand, & Tummolini, Luca. (2018). Varieties of abstract concepts: development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170121. doi:doi:10.1098/rstb.2017.0121

Borghi, Anna M, & Binkofski, Ferdinand. (2014). *Words as social tools: An embodied view on abstract concepts* (Vol. 2): Springer.

Borghi, Anna M, Binkofski, Ferdinand, Castelfranchi, Cristiano, Cimatti, Felice, Scorolli, Claudia, & Tummolini, Luca. (2017). The challenge of abstract concepts. *Psychol Bull, 143*(3), 263-292. doi:10.1037/bul0000089

Bozeat, Sasha, Lambon Ralph, Matthew A., Patterson, Karalyn, Garrard, Peter, & Hodges, John R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38*(9), 1207-1215. doi:https://doi.org/10.1016/S0028-3932(00)00034-8

Brysbaert, Marc, Warriner, Amy Beth, & Kuperman, Victor. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911. doi:10.3758/s13428-013-0403-5

Buckner, Cameron. (2018). Empiricism without magic: transformational abstraction in deep convolutional neural networks. *Synthese, 195*(12), 5339-5372. doi:10.1007/s11229-018-01949-1

Cortese, Michael J, & Fugett, April. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384-387.

Davis, Charles P., Altmann, Gerry T. M., & Yee, Eiling. (2020). Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology, 37*(1-2), 142-153. doi:10.1080/02643294.2019.1710124

Dove, Guy. (2018). Language as a disruptive technology: abstract concepts, embodiment and the flexible mind. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170135.

Dove, Guy, Barca, Laura, Tummolini, Luca, & Borghi, Anna M. (2020). Words have a weight: language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research*. doi:10.1007/s00426-020-01438-6

Fellbaum, C. (Ed.) (1998). *WN: An Electronic Lexical Database*. Cambridge: MIT Press.

Juhasz, Barbara J., & Yap, Melvin J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods, 45*(1), 160-168. doi:10.3758/s13428-012-0242-9

Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., & Baciu, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: a behavioral and fMRI study. *NeuroImage, 44*(3), 1152-1162. doi:10.1016/j.neuroimage.2008.09.043

Kounios, John, & Holcomb, Phillip J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 804-823. doi:10.1037/0278-7393.20.4.804

Langland-Hassan, Peter, Faries, Frank R., Gatyas, Maxwell, Dietz, Aimee, & Richardson, Michael J. (2021). Assessing abstract thought and its relation to language with a new nonverbal paradigm: Evidence from aphasia. *Cognition, 211*, 104622. doi:https://doi.org/10.1016/j.cognition.2021.104622

Langland-Hassan, Peter, Gauker, Christopher, Richardson, Michael J., Dietz, Aimee, & Faries, Frank R. (2017). Metacognitive deficits in categorization tasks in a population with impaired inner speech. *Acta Psychologica, 181*, 62-74. doi:https://doi.org/10.1016/j.actpsy.2017.10.004

Löhr, Guido. (2022). What Are Abstract Concepts? On Lexical Ambiguity and Concreteness Ratings. *Review of Philosophy and Psychology, 13*(3), 549-566. doi:10.1007/s13164-021-00542-9

Markman, Ellen M. (1981). Two different principles of conceptual organization. In M.E. Lamb & A.L. Brown (Eds.), *Advances in Developmental Psychology* (pp. 199-236). Hillsdale, NJ: Erlsbaum.

Mirman, Daniel, & Graziano, Kristen M. (2012). Individual differences in the strength of taxonomic versus thematic relations. *Journal of Experimental Psychology: General, 141*(4), 601-609. doi:10.1037/a0026451

Penn, Derek C, & Povinelli, Daniel J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a theory of mind. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1480), 731-744. doi:doi:10.1098/rstb.2006.2023

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Yee, Eiling. (2019). Abstraction and concepts: when, how, where, what and why? *Language, Cognition and Neuroscience, 34*(10), 1257-1265. doi:10.1080/23273798.2019.1660797

Yee, Eiling, Jones, Michael N, & McRae, Ken. (2018). Semantic memory. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, 3*, 1-38.