

Millikan's Consistency Testers and the Cultural Evolution of Concepts

Nicholas Shea

Abstract

Ruth Millikan has hypothesised that human cognition contains 'consistency testers'. Consistency testers check whether different judgements a thinker makes about the same subject matter agree or conflict. Millikan's suggestion is that, where the same concept has been applied to the world via two routes, and the two judgements that result are found to be inconsistent, that makes the thinker less inclined to apply those concepts in those ways in the future.

If human cognition does indeed include such a capacity, its operation will be an important determinant of how people use concepts. It will have a major impact on which concepts they deploy and which means of application (conceptions) they rely on. Since consistency testers are a selection mechanism at the heart of conceptual thinking, they would be crucial to understanding how concepts are selected – why some are retained and proliferate and others die out. Hence, whether consistency testers for concepts exist, and how they operate, is an important question for those seeking to understand the cultural evolution of concepts, and of the words we use to express them.

Word count: 10,080

Sections

- (1) Introduction
- (2) Millikan's Hypothesis
- (3) Consistency Testers Characterised
- (4) Metacognitive Role
- (5) Objections
- (6) Evidence
- (7) Role in the Evolution of Concepts and Language
- (8) Conclusion

(1) Introduction

Ruth Millikan has argued that human cognition includes a capacity for checking the soundness of our concepts. These 'consistency testers' check whether different ways of applying a concept agree in their verdicts. Concepts are applied directly on the basis of perception, indirectly by inference from other properties, and via relying on the testimony of others ('it's raining'). When a concept is applied in two different

ways, for example when I recognise a person by sight and by their voice, the consistency tester checks that the properties I ascribe from the two different perspectives agree. If I judge Fa from one perspective and $\neg Fa$ from another perspective, then something has gone wrong.

This is not a matter of making explicit judgements about our thoughts, but is instead supposed to be an automatic and unreflective process that monitors our cognition as it unfolds. Where a contradiction arises, we become less inclined to apply the concept in those ways: we revise them or stop using them. If a concept generates contradictions repeatedly, we are inclined to give up using it entirely. Or so Millikan claims.

The suggestion that cognition contains consistency testers is a powerful idea. It would tell us much about how people acquire and use concepts. It would also be a key driver of the cultural evolution of our conceptual repertoire, and of the linguistic terms we use to express our concepts. However, Millikan's hypothesis has received little attention to date. The purpose of this paper is to show that the hypothesis ought to be considered seriously. I will set out the idea of consistency testers, assess their promise and some objections, and consider their evidential support.

If confirmed empirically, the way this form of monitoring impacts our dispositions to use concepts would make it a form of procedural metacognition. Consistency testers will be a key part of the explanation of why people adopt and use some concepts, and reject others. This in turn will help explain why some concepts proliferate and spread culturally when others do not.

I start by setting out Millikan's hypothesis (section 2), characterising the consistency testing process (section 3) and showing that the way they are supposed to work is metacognitive (section 4). I go on to consider some objections raised in the literature (section 5) and to assess whether there is evidential support for the hypothesis (section 6). Finally, I say something about the role consistency testers are likely to play, if the hypothesis is confirmed, in the cultural evolution of concepts and of language (section 7).

(2) Millikan's Hypothesis

Millikan introduces consistency testers to play a central role in her theory of content. A theory of content is an account of how representations get their meaning, why they refer to or are about things in the world. A theory of content for the honeybee nectar dance will say what makes it the case that dances are about the location of nectar. Crucially this will show why a certain dance is correct when there is nectar at a certain location and incorrect otherwise. A theory of content for concepts will say what makes it the case that our concepts refer as they do: that my concept **DOG** picks out dogs, **SMOOTH** the texture smoothness, and **MUM** my mother.

In *Language, Thought and Other Biological Categories* (1984, 'LTOBC') Millikan advanced a radical new theory of content. Her treatment is detailed and comprehensive. Consistency testers are just one part of a very rich theory, other parts of which have been extensively debated in the literature. One distinctive feature of Millikan's theory is the need for a 'consumer' of a representation to fix its content. Theorists had previously tended instead to start with the way a representation is produced. A representation that is exquisitely sensitive to the presence and orientation of a line in the visual field is often thereby taken to be about line orientation. Millikan instead looks first to the way a representation is used. If I want to work out what 'ugali' means in an unfamiliar country, I see what people get when they order 'ugali' in a café. I read off the meaning of the word from the pattern of behaviour it prompts. Millikan's idea is similar: the meaning of mental representation R is constituted by the way some consumer system reacts to it, by the kind of behaviour it prompts. The behaviour that the consumer produces in response to R effectively takes for granted that a certain condition C obtains. C is then the content of R.

For some representations there are consumer systems that produce overt behaviour. The consumer bees who observe a nectar dance fly off a certain distance in a certain direction before searching for nectar. The cat's internal representation of the presence of a mouse prompts a suite of stalking and catching behaviours. Our visual representation of the size and shape of an object conditions the grip formed by the hand on the way to grasping it. For Millikan, content is fixed at the level of a complete representation, like a judgement, that has a truth condition or satisfaction condition. Concepts are components of judgements. The content of a concept is determined by the way judgements that contain the concept are used. Some judgements issue directly in behaviour. For example, my judgement that *mangoes are tasty* might lead me to select a mango from the supermarket shelf. For many concepts, however, there do not seem to be consumer systems issuing directly in behaviour that consume judgements containing them. Since Millikan wants content to be fixed by representation consumers, she needs an account of how these representations are consumed.

For Millikan, consumer systems don't just fix content. The way a consumer system reacts to a representation is also a test of whether the representation has been applied to the world appropriately – of whether the system is getting it right. There are, then, three reasons why many concepts are not consumed by systems that issue directly in action (1984, 142). First, it would be too dangerous. The consequences of relying on a concept that is misapplied to the world, or is confused, could be death or injury. A cognitive checking process is less costly. Second, Millikan argues that actions are always based on large number of representations at once – background assumptions and representations of what is currently the case. When an action goes wrong, it would be hard to find the culprit. Millikan wants a mechanism that can test a few concepts at once. Third, we have too many concepts, and acquire them too quickly, to be able to test them all in action.

This is where consistency testers come in. The idea is that concepts are tested not in action but in cognition. Millikan's hypothesis is that humans have a capacity for checking different concept-producing mechanisms for consistency with one another: 'the law of noncontradiction is, in the first instance, a *concept tester*' (1984, 143). When I hear a bird and categorise it as a sparrow, does that agree with the way I apply the concept SPARROW when I see one? When a surface looks smooth, does it also feel smooth to the touch? The basic test is that when I judge Fa from one point of view, I should not reach a contrary judgement from another point of view. If I do, the contradiction is a sign that either or both concepts are defective (the concept I used to pick out F or the concept I used to pick out a). The consumer system that serves to fix the content of a judgement or other inner sentence S (and hence of the concepts it deploys), 'is the consistency tester qua tester of *other* programs that can produce tokens of the same type as S or negations of these' (1984, 146).

That is the innovation I want to examine in this paper. Before going into details, I should say more about how consistency testers have figured in Millikan's work. All the key features of consistency testers were presented right at the start, in *LTOBC*. There was so much in *LTOBC* that it is not surprising that some parts have been overlooked. Millikan put forward consistency testers somewhat tentatively, since the hypothesis was not based on experimental evidence but just on commonsense reflection. Plus, they are not essential to her theory of content. There are representations whose consumers produce behavioural outputs (like the bee dance), so discussion has tended to focus on those. Consistency testers are a specific, optional hypothesis about how concept-involving representations are in fact consumed (Godfrey-Smith 2013). As Millikan says, concepts could get their content by having a direct role in guiding behaviour (1984, 127-130). Consistency testers are not then something on which the viability of the whole teleosemantic approach to content depends. They are however a centrepiece of *LTOBC*'s argument against individuating concepts by conceptions ('meaning rationalism') – conceptions being beliefs or means of application of a concept (or, for other theorists, neo-Fregean senses).

Consistency testers did not feature in Millikan's papers in the aftermath of *LTOBC* (1984). In the widely-cited summary published five years later (Millikan 1989), they do not figure at all. They are back with full force in Millikan's brilliant theory of empirical concepts, *On Clear and Confused Ideas* (Millikan 2000). In one sense, that is what the book is about: the ability to successfully reapply the same concept from a range of different perspectives. It is advertised on the cover as, 'the first in-depth discussion on the psychological act of reidentification'. But in *On Clear*, teleosemantics takes a back seat. 'In [*On Clear and Confused Ideas*] I tried to get on as much as possible without reference to selection processes, not because I no longer thought them central, but in a climate that was so hostile to "teleosemantics" it seemed politic' (Millikan 2017, 8). In her most recent book Millikan does more to explain how her theory of empirical concepts is integrated with her teleosemantic approach to content (Millikan 2017).

(3) Consistency Testers Characterised

I will focus on the account of consistency testers put forward in *LTOBC* and then mention a few refinements from later work. Millikan's hypothesis is that we have a capacity for testing our judgements for consistency. This is a way of checking that different representation-producing mechanisms agree with one another. An empirical concept can be applied in a variety of ways: through a variety of perceptual routes; by inference from other beliefs; and based on the testimony of others. Do these different means of identification pick out the same thing? One way to test that is to see whether they agree with each other in judgement. Millikan's idea is that agreement in judgement is a test for the reliability with which a concept is applied. If you judge Fa from one perspective and at the same time judge $\neg Fa$ from another, that is a sign that something is wrong with the way you are disposed to pick out F , or a , or both.

Millikan argues that we recognise a contradiction by appreciating that some attributes exclude one another. If I recognise my friend Hemara in part because she is tall, and I then see what I take to be Hemara and judge that she is short, something has gone wrong. A person cannot be both short and tall at the same time. Properties come in groups that exclude one another on the ground of a common substance. *Tall* and *short* exclude one another as applied to the substance that is an individual person. The contradiction between my judgements Fa and Ga , when F and G are contraries, shows that something has gone wrong with some of the concepts involved. Judging that Hemara is tall and that Hemara is clever does not present a contradiction. Millikan's consistency testers rely on the subject recognising which properties exclude one another. She also requires the subject to have the concept of negation. For Millikan, the primary function of negation is to act on the predicative aspect of a sentence, so that the negated sentence represents a positive state of affairs (1984, 221-9). For example, 'Hemara is not tall' asserts that Hemara's height is amongst the heights that are consistent with not being tall. The concept of negation performs a similar function in thought. The consistency tester registers that something has gone wrong when we judge Fa and $\neg Fa$ at the same time.

Consistency testers are way of testing concepts-in-judgements against one another piecemeal, rather than testing one's whole conceptual scheme at once. Millikan's innovation was to advocate consistency testing, not as a way of checking that judgements are true, but as a way of checking that concepts are univocal. *Beyond Concepts* (2017) emphasises that what is mainly being tested are different ways of applying a concept, different ways of tracking its referent. (There, ways of applying a concept are called 'unitrackers', and concepts are replaced with 'unicepts', the change of terminology serving to emphasise that they are not shared between different individuals.) Where two ways that a thinker applies a concept produce contrary judgements, there is probably a fault with one of the ways in which the concept was applied.¹ That does not pin down the culprit uniquely. The subject is applying a substance concept (e.g. of an individual, Hemara) and two attribute concepts (tall, short), and is assuming that the attributes are contraries (short \rightarrow \neg tall). Any or all of

¹ There could instead be an error due to noise – even highly reliable ways of applying a concept are not perfect.

these could be mistaken. But Millikan argues that this is a considerable improvement over standard coherentist views, which see the whole body of the thinker's beliefs being tested at once. Inconsistency could rebound on all the concepts involved in the pair of inconsistent judgements, or it could be that factors about the individual case make the thinker more inclined to doubt one or other way of applying the concept. For example, although I knew Hemara well from video meetings, when I then meet her in person I readily give up my former conception of her as tall.

Concepts can also be applied on the basis of inference from other concepts. Suppose I carry out various tests on a colourless volatile liquid, including measuring its boiling point, and infer from these properties that it is ethanol. In fact the thermometer was over-reading and the substance was actually methanol. The fault there lies not with my conception of the properties of ethanol but with how I was measuring temperature. It is not so easy to pin down the source of contradictions in these kinds of cases. Since much of our conceptual thought occurs when we infer from some thoughts to others, the groups of concepts being tested by a given consistency test are probably often rather larger than Millikan suggests.

Consistency testing is postulated as a monitoring mechanism. It is also supposed to have an impact on subsequent cognition. Where a consistency check encounters a contradiction, that makes the thinker less inclined to use the concepts involved and less inclined to apply the concepts in those ways. This hypothesis about the downstream effects of encountering inconsistency is an integral part of Millikan's account of consistency testing. It is a substantial empirical claim. If it turns out to be true, it would be an extremely important feature of the way human conceptual thought functions, forming a key part of the way we acquire and retain concepts, with implications for the cultural evolution of concepts and of the words we use to express them.

Many of our concepts are initially acquired through language. I can read about 'monotremes' in a book and learn some things about them (1984, 152). That gives me a concept, MONOTREME (an 'inner term'), and some ways to apply it (using the facts I have learned). For Millikan, recognising a monotreme through the testimony of others ('that's a monotreme'), or through applying the facts I have learned about it, is just like recognising it perceptually. These are all just means of applying my MONOTREME concept. Our inner consistency tester checks them against one another. Concepts that survive this test continue in use (1984, 145, 302). Concepts the prompt contradiction and confusion die out. Consistency testing is the internal selection mechanism at the heart of the way concepts proliferate (2017, 80). Internal selection operating over reproductively established families of inner terms is supposed to confer evolutionary functions on them (1984, 145). It also accounts for the cultural selection of linguistic terms. A name survives because speakers are using it and hearers are understanding it in ways that produce agreement in judgement (2017, 99). For Millikan, then, consistency testers play a key role in the cultural evolution of language and of concepts.

Consistency testing is also playing an epistemic role for the thinker. It is how we can tell that our concepts are clear and not confused (1984, 142-145). When the various ways in which a concept is applied agree with one another in judgement, that is a sign that we are correctly applying the concept to the world and that there is a common referent to these uses – for example, that they all pick out organisms belonging to the same objective biological class. The common referent is the underlying reason that, even when we pick out monotremes in very different ways (visually, by description, through testimony), the things we pick out display a consistent set of properties (being warm blooded, laying eggs, etc.). It is because the natural kind *monotreme* has a variety of stable properties that our judgements agree with each other when we apply the concept MONOTREME in different ways. Consistency testers are checking that. They are telling us which ways of applying a concept pick out something objective (2017, 80). Passing the test also tells us that the different ways in which a concept is applied are consistent with one another – that they pick out the same thing.

In *On Clear and Confused Ideas* (2000) Millikan says more about how empirical concepts are tested epistemically. She enumerates four ways the thinker can test whether their thoughts are confused or clear. As before, consistent convergence in independent methods of judgement is evidence that the concepts involved are being applied to the same thing (that these judgments have an objective, univocal source: 103). Second, a concept that picks out a real objective substance (individual, kind, etc.) should be applicable in many different ways. A lack of variety in the ways a concept can be applied is a sign that it may be empty. Third, in some cases concept application is consistent only when applied from a certain perspective, agreeing in judgement only within a given context. For example, I might have a concept of an actor and musician called 'John Baker'. As I learn more about him (actually, them), I discover that he has some quite inconsistent views. The songs are sensitive, introspective and highly political, whereas the actor comes across in his public appearances as bombastic, party-loving and completely apolitical. It dawns on me that these are two different people. If contradictions are systematically correlated with across-context judgements, that is a sign that the concept is equivocal, picking out two different referents at once. Fourth, if two different concepts are applied on the basis of the same properties, and are not differentiated by contrary properties, that is a sign of redundancy: that the thinker may have two different concepts of the same referent. (This is how Lois Lane could come to realise that Superman is Clark Kent.) *On Clear's* theory of concept testing thus extends widely, but consistency testers still play a prominent role (in the first and third tests).

Consistency testers were postulated by Millikan in *LTOBC* to play the consumer role for concept-involving judgement. When I learn about monotremes, my new MONOTREME concept has never led to any overt behaviour. Nevertheless, it has been input into a consumer system, for Millikan – the consistency tester. The consumer of MONOTREME is the process of testing inner sentences containing the concept MONOTREME against one another (1984, 146). Millikan's claim about content is that the

content of the concept is fixed by the evolutionarily normal conditions for the operation of the consistency testing mechanism. For the concept MONOTREME, the normal condition is that it should be applied to monotremes, or carry correlational information about monotremes (1984, 146), or that inner sentences containing the concept MONOTREME consistently map onto states of affairs involving monotremes (1984, 301-2). Or so Millikan argues.

There is much more to be said about Millikan's theory of content constitution. One worry is whether relying on normal conditions for agreement in judgement is sufficient to deliver a determinate content for the concepts involved. Situations in which a judgement is applied to the world share many characteristics. Which of these objects and attributes get to enter into the content of the judgement? Where consumer systems issue in action, teleosemantics appeals to the way that action unfolds, and its effect on the organism, e.g. via selection, to home in on determinate contents. It is unclear whether internal comparison of two applications of a concept in judgement has as good resources for dealing with indeterminacy. When writing after LTOBC, it often seems that Millikan has something input-driven in mind when talking about the thing to which a concept is applied (2000) or the source of the information that is tracked by a unitracker and brought together in a unicept (2017). Information in the bare correlational sense is notoriously liberal. It seems that Millikan has in mind a kind of triangulation that cuts down on the liberality, for example, 'The referent is determined as being the actual subject of the information that has been gathered and linked together' (Millikan 2022, 14). Substance templates also enter the picture and play a role in constraining what kind of thing a concept can refer to (2000).

A full account doubtless relies on all these elements, but there is not space to explore these issues here. This paper is not about how content is metaphysically constituted. My interest is in the empirical hypothesis that humans have consistency testers, what that would tell us about how cognition works, and its implications for the cultural evolution of concepts and language.

As a hypothesis about our cognitive mechanisms, consistency testing promises to be a crucial part of unsupervised concept learning. Computational models of representation learning divide into supervised and unsupervised learning algorithms. In supervised learning, outputs of the system are assessed for correctness. When the system responds to an input with an incorrect output, that is used as a teaching signal to change the configuration of the system so as to produce a more accurate output the next time that input is encountered. Supervised learning in deep neural networks has proven to be extremely powerful. For example, it can produce better-than-human performance on tasks like image classification (Krizhevsky et al., 2012; Eslami et al., 2018), natural language processing (Bahdanau et al., 2014; Brown et al., 2020; Floridi & Chiriatti, 2020) and protein folding (Jumper, et al. 2021). Unsupervised learning looks for patterns in the training data without relying on any overt teaching signal. While unsupervised learning has always been important in machine learning research, the successful recent neural network models have almost all relied on supervised learning. Turning to psychology and cognitive neuroscience, however, it is clear that a

considerable amount of unsupervised learning must take place in human cognitive mechanisms.

Millikan's theory offers us a picture of how unsupervised concept learning could work. We pick up information from the environment and index it to internal representations (mental words). This gives us a way of applying the representation (concept) to situations registered at input. We start re-using the concept in situations like the ones we have observed. There need be no external feedback. But the system is constrained by consistency testing. The system needs to be able to recognise that different tokens of the same mental word are tokens of the same concept, and that some predicates are contraries ($Gx \rightarrow \neg Fx$) (1984, 146). It expects new situations in which it applies the concept to have the same properties as observed in the past. When that fails, the means of application are altered or the concept is retired from use. This is a substantial hypothesis about how unsupervised concept learning proceeds in human cognition.

(4) Metacognitive Role

There is clearly something metacognitive about consistency testing. But it is not a matter of making judgements about one's thoughts. The thinker is sensitive to whether their judgements are consistent or inconsistent without explicitly reflecting on their judgements or forming a meta-level judgement to that effect. Consistency testing does not call for the capacity to self-ascribe thoughts (2000, 102). Nor do we need to appreciate how we are applying a concept (1984, 322). We may find ourselves disposed to categorise something as an X while having no conception of why we are so-disposed. Think of the way that you recognise familiar people. You simply recognise that this is so-and-so. When we do apply a concept by inferring from other facts (this substance is ethanol because it is a clear liquid, boils at 78 degrees, etc.), then we can reflect on and correct our beliefs about how to apply the concept. But the consistency test does not depend on that ability (2000, 105). It applies to the judgements we form irrespective of how we apply the concept.

Nevertheless, consistency testing is clearly metacognitive. It involves both monitoring and control (Nelson and Narens 1990). It applies to the thinker's own cognitive processes and has an impact on how they are carried out thereafter. Since it does not involve explicit beliefs or judgements, it is a form of procedural rather than analytic metacognition (Proust 2012b, 2013a, 293-294).

Millikan does not treat consistency testing as an unconscious process. The processes that test for consistency need not be conscious, but Millikan seems to assume that the deliverances of the consistency tester are consciously available to the thinker (2000, 103, 105). They offer the thinker a way of appreciating whether their concepts are confused or clear. This suggests that recognising consistency or inconsistency is accompanied by an epistemic feeling (although Millikan does not put it in those terms). Contradiction in judgement is apparent to the thinker and that is why their dispositions to apply the concept change. Since explicit judgements need

not be involved, epistemic feelings are suited to playing this role. They result from consistency monitoring and affect how the concepts are applied in the future. (Epistemic feelings may have non-conceptual content – a correctness condition and/or satisfaction condition – as a result of their role in cognition, but that is an additional, optional theoretical commitment.)

The suggestion that consistency testing generates epistemic feelings would help underpin another of Millikan's claims. She argues that consistency testing is how a thinker checks that they know what they mean in making a judgement ('grasp of meaning': 1984, 322), that they know what they are thinking of (2000, 95). This knowledge is a matter of being able successfully to reidentify the referent of a concept from a variety of perspectives (2000, 91-94, 95-105). Consistency testers are the way the thinker checks that they can do so, as we have seen. A positive verdict from consistency testing thus delivers a signal to the thinker that they have a grasp of meaning. An epistemic feeling would show up in the thinker's mental life as a signal to that effect. Knowing what one is thinking would, then, not just be a matter of the thinker in fact being able to apply a concept correctly from a variety of perspectives. The thinker is also able to register that fact. The deliverance of the consistency checking mechanism is an internal epistemic signal, available to the thinker, which plays a role in regulating how their thinking unfolds [cp. reference in submission].

(5) Objections

Having set out Millikan's hypothesis about consistency testers and explained the role it plays in her theorising, I now turn to assessing her claim. Although it has received much less attention than other parts of her theory, Millikan's consistency testers have attracted some comment. Godfrey-Smith (2013) characterises the hypothesis as a 'strong' empirical claim and emphasises that its truth is not required for Millikan's theory of content to be sustainable. Objections have been raised by Robert Rupert (1999) and Manolo Martinez (2013). I will consider those now, before going on in the next section to consider the empirical evidence.

Rupert (1999) claims that there is a circularity in the way that Millikan specifies the operation of consistency testers. He argues that this circularity undermines the content-determining role which Millikan wants consistency testers to play. Although we are not here concerned with the metaphysics of content, it would be a problem if Millikan has failed satisfactorily to specify what a consistency tester is.

The alleged circularity arises from the very notion of consistency. To successfully check for consistency is to successfully judge Fa from two different perspectives. Such success depends on the concepts involved having content. The checking process is successful if it correctly concludes that the two judgements have the same, hence non-contradictory content. However, Rupert points out, past successes in the tester's performing its proper function are supposed to make it the case that the judgements, hence the concepts involved, have content. How can the

mechanism be testing for consistency if it is consistency that makes it the case that the concepts involved have content?

The answer is supposed to be that the operation of the consistency tester both makes it the case that representations have content and makes it the case that what it is doing is testing for consistency. It is the same kind of move that is familiar from teleosemantics applied to over behaviour. Past occasions when a representation prompts behaviour that is stabilised both make it the case that the representation counts as correct and make it the case that the representation has content. There are legitimate questions about whether this move, applied to consistency testers, is capable of delivering determinate content. But does it undermine Millikan's way of specifying what a consistency tester is?

It does not, because Millikan explains what a consistency tester is, not in terms of contents, but in terms of operations on vehicles. It is a mechanism that can recognise multiple tokens of the same representation vehicle (concept-involving judgement) and can recognise predicate negation. It recognises an inconsistency by recognising that concept-application mechanisms have issued in the judgement Fa and also the judgement $\neg Fa$. All of that can be specified in terms of how the mechanism operates on representational vehicles. Claims about the evolutionary function of this mechanism are important to the theory of content, but they are not essential to being able to individuate the mechanism. We can say what it is to be a consistency tester without presupposing that it does indeed test for semantic consistency. (We could call it a syntactic consistency tester or a putative consistency tester.)

Martinez (2013) raises a related concern. He argues that there is no fact of the matter about whether a certain mechanism is a consistency tester (62). He offers Millikan a different answer to the just-mentioned concern about content determination. Rather than fixing facts about consistency and content all at once, a mechanism that operates syntactically in the way Millikan hypothesises could function to test concepts for consistency whose content has already been fixed. This fits with Millikan's contention that some concepts are tested directly in action (2000, 108). Concepts that have been used to reidentify individuals, kinds or stuffs in order to act on them are thereby tested for adequacy without relying on consistency testers. They provide 'a certain sort of foundationalist base for the conceptual abilities later employed in theoretical knowing' (108). New concepts could then be tested for consistency against this store of action-tested concepts (Martinez 2013, 62).

Martinez's concern is that Millikan does not have the resources to say whether a mechanism is testing for consistency, or the closely-related notion of consistency*, where consistency* is a matter of consistency in judgements about states of affairs with which the thinker has some kind of causal contact. That is a legitimate concern about how a consistency-testing mechanism can determine content – how it can determine verification-transcendent content, content concerning states of affairs with which the thinker makes no causal contact. But Martinez's 'no fact of the matter' claim does not undermine the way I have identified consistency testers here. For our

purposes, a consistency tester is a mechanism that carries out certain syntactic operations on internal representations; and which, when it encounters a syntactic contradiction, has certain effects on downstream processing. Martinez's point is not an objection to the hypothesis that there is a mechanism which plays that role in human cognition.

In short, although Rupert and Martinez have raised important objections to the role of consistency testers in a theory of content, their points do not undermine the way consistency testers are individuated as a psychological mechanism. But what is the evidence that the hypothesised mechanism exists? That is the question we turn to next.

(6) Evidence

Is there any evidence for the existence of consistency testers? The first place to turn is to the preeminent philosophical theorist of metacognition, Joëlle Proust. In a series of papers and a compendious book (2013a), Proust offers a comprehensive account of all the kinds of metacognition that are at work in human cognition. Do we find consistency testers in her collection?

Proust shows that epistemic feelings are generated by cognitive processes of many different kinds. Proust's overall framework is that epistemic feelings are calibrated based on the history of previous results of performing mental actions of a given type (Proust 2008). This provides a way of making metacognitive monitoring reliable. Every mental activity has a 'viability core': a range of circumstances in which it can operate reliably. The thinker may not know what the viability core of a given mental activity is. They rely on feedback from the results of performing that activity to sense where the limits of the viability core lie (2013b, 199-206). For example, a thinker can come to learn in which circumstances they are likely to get it right when they attempt to retrieve a person's name from memory, and in which circumstances their memory is likely to be unreliable. If given false feedback, the feeling of confidence accompanying memory retrieval would become a poor guide to the accuracy of memory. Since we have access to some accurate feedback about whether we were operating within the viability core of the relevant process – that we got the name right – our epistemic sentiments are trained to be at least somewhat reliable.

Applying this framework to our case, Proust takes it (uncontroversially) that judging is a mental action whose aim is to produce true beliefs (Proust 2012a, 2013a, 172). Mental acts like judging are sensitive to two norms: accuracy and coherence (Proust 2013b). Coherence is a broader notion, but Proust does see a role specifically for consistency in affecting confidence. She endorses Asher Koriat's self-consistency model in which agreement with oneself in judgements increases confidence and conflict decreases confidence (Koriat 2012a, b). Koriat's model concerns confidence in judgements (or semantic memory, etc.), but it does support Millikan's hypothesis that there is a cognitive mechanism for testing consistency between judgements. Proust's theory is based on philosophical reflection, while also drawing on empirical results.

The fact that she independently concludes that judgements are monitored for coherence and consistency lends some support to Millikan's hypothesis that there are mechanisms for testing consistency between judgements, if not for Millikan's distinctive claim about this being a test for the univocity of concepts, and thus its having an impact on downstream concept use

Sperber, et al. (2010) postulate a consistency testing mechanism as part of their account of 'epistemic vigilance' – of the way we check what we are told by others for its veracity. Although their focus is testimony, their claims about consistency and coherence checking are supposed to be general properties of cognition. They ask, 'what happens when the result of processing a new piece of information in the context of existing beliefs is a contradiction?' (375). The answer is that the thinker comes to distrust one of the sources of information (testimony, memory, etc.). The focus here is confidence in judgement, but Sperber et al. are assuming that there is a mechanism which checks mental states for consistency. (As well as explicit contradictions, it also registers incoherence, i.e. incompatibility with background beliefs: 375.) Finding an inconsistency triggers a procedure dedicated to its assessment (376). Sperber et al. do not share Millikan's more specific claims about the downstream effects on concept application (distrusting a source is one way of becoming less inclined to rely on a means of application, but it has much a wider effect). Nevertheless, what they say is compatible with Millikan's proposal. Since Sperber et al.'s account is supported by a wide range of evidence (largely linguistic), as well as theoretical reflection, it too supplies some support for the consistency testing hypothesis, although not for the particular claims that Millikan makes about the impact of inconsistency on concept use.

What about experimental evidence from psychology and cognitive science? Behavioural experiments on metacognition provide general support for the idea that there is a mechanism that registers when occurrent beliefs or judgements are inconsistent. Koriat's self-consistency model, mentioned above, appeals to consistency in beliefs to explain confidence reports (Koriat 2012a). An important line of evidence in support of Koriat's model is the correlation between within-person response consistency and confidence. Participants are asked to choose an answer to a question on several occasions. Answers that are given consistently in answer to a given question attract higher confidence than those which vary. This effect is found for a range of tasks: general knowledge questions, word matching, social beliefs and attitudes, and perceptual comparisons. Confidence tracks consistency better than accuracy. This is good evidence that cognition systematically monitors consistency between beliefs, as Millikan claims. On the other hand, there is evidence that people can be surprisingly tolerant of inconsistency where the judgements are made in different different contexts (Astuti and Harris 2008, Legare and Shtulman 2018). This fits with Millikan's hypothesis that consistency checking operates, not on a whole person's store of beliefs, but on occurrent judgements at the time at which they are made.

More broadly, many neural processes are concerned with the expectedness or unexpectedness of stimuli or the elicitation of surprise in cognition (Polich 2007, Kutas

and Federmeier 2011). Such signals are found in the context of categorization, differentiating between expected and unexpected categorization (Hamm, et al. 2002). This supports Millikan's contention that cognitive processes of comparison are at work when people are applying concepts to the world. But it suggests that the comparison processes are much wider than simply checking for inconsistency between judgements of the form Fa and $\neg Fa$.

Predictive coding has been proposed as a ubiquitous computational principle in the brain (Friston and Stephan 2007). It is doubtless a feature of some cognitive processes. This suggests that it is plausible that a new judgement should be checked against a previous judgement (an expectation), with any inconsistency generating an error signal. In predictive coding models, the error signal has downstream consequences in that it changes the thinker's expectations. However, the specific consequences predicted by Millikan's hypothesis, namely a disposition to refrain from using the concepts involved or to rely on those modes of application, is not a standard consequence of predictive coding models.

The literature on neural signals of incongruence and error is enormous and impossible to summarise here. An overall impression is that there are a large variety of error monitoring mechanisms, many of which could have an effect on how concepts are deployed. Millikan's mechanism relies on an explicit concept of negation. However, it might simply be enough to apply two predicates which the thinker appreciates are contraries, Fa and Ga , without the thinker having to form the negative judgement $\neg Fa$. If the thinker appreciates, even tacitly, that F and G are drawn from a range of predicates that exclude one another, just judging Ga , having previously judged Fa of the same individual, might show up as unexpected and thus be enough to generate an error signal. It is a very general feature of cognition that psychological processes generate expectations and that error signals arise when those expectations are not met. (One does not have to subscribe to predictive coding as an all-encompassing theory of the mind to accept the prevalence of expectations and error-related signals.) So there may be a whole suite of mechanisms, including consistency testers in Millikan's sense but extending much more widely, which generate epistemic feelings when concepts are relied on in ways that produce unexpected consequences. These mechanisms seem to produce graded error signals, based on how unexpected the consequence was, rather than a binary consistent / inconsistent verdict.

What of the downstream effects of detected inconsistency? Very little research has been carried out that speaks to Millikan's specific hypothesis about the downstream effects of inconsistency on concepts. The issue seems to be empirically tractable but it has not yet, so far as I know, been tested directly. Confidence is, however, in general related to behaviour in the ways one might expect. Where the first answer to a question is given with high confidence, it is more likely that the same answer will be given in the future (Koriat 2012b). People are inclined to bet on their responses in line with their reported confidence (Koriat 2011). Koriat's data show an effect of inconsistency on confidence in beliefs, but that does not tell us whether there is an effect on how people would apply and rely on the concepts involved. So far as I know, there has been no direct experimental test of Millikan's hypothesis about the

effect of inconsistency on subsequent concept use. As Koriat remarks, while there has been a lot of work on how metacognitive monitoring operates, there has been much less research on the way the output of monitoring processes feeds into the regulation of cognitive processes and behaviour (Koriat 2012c).

A converging line of evidence comes from work on judgement and decision making, particularly from those adopting a dual systems approach. The cognitive reflection test (CRT) is one example. The CRT consists of a series of problems for which there is an intuitive but incorrect answer that clashes with the correct answer, which could be worked out more carefully.² People are asked to report a ‘feeling of rightness’ after giving their initial answer. The feeling of rightness is lower for conflict problems and predicts the amount of time people spend re-thinking their answers (Thompson and Johnson 2014). People can also be asked, after they have finished solving a task, whether they think they have made an error (Fernandez Cruz, et al. 2016). These ‘feelings of error’ have been elicited for both the CRT and a mental mathematical calculation (Ackerman and Thompson 2017). The feeling of error also reflects inaccuracy, but it is not clear from the data gathered to date what other effects it has on downstream cognition, let alone any specific effects on concept use.

One way to test an aspect of Millikan’s hypothesis would be to use a ‘cue utilization’ framework (Koriat 1997). Having first ascertained which cues a subject is relying on in making a judgement, they would then be presented with evidence that invites a contradictory judgement. Millikan’s hypothesis predicts that they should thereafter be less inclined to apply the concepts involved in the ways that they did. So the cues they use in applying the concepts should have changed (or the weights attached to them).

There have been a few experiments specifically investigating how metacognitive processes apply to concepts. One approach proceeds by teaching people new concepts, either of natural categories (e.g. types of birds) or artificial categories (e.g. cartoon bugs). For instance, researchers have trained participants to classify items into unfamiliar categories (e.g. artists’ paintings) and then asked them to make *judgements of learning* about each item during training, and confidence judgements during test (Kornell and Bjork 2008, Kornell, et al. 2010). In similar research, Jacoby, et al. (2010) introduced a further measure, the *category learning judgement* (CLJ), a judgement of how well one has learned to categorize items under the concept one has just learned.

CLJs show the broad pattern displayed by other kinds of metacognitive judgement: they are generally reliable, both in terms of bias and calibration, when using both natural concepts (Jacoby, et al. 2010) and declarative concepts (Rawson, et al. 2015) in the lab, and when teaching students statistical concepts in the classroom (Hartwig and Dunlosky 2017). CLJs can be readily affected by factors that are only very indirect proxies for reliability, such as repetition (Wahlheim, et al. 2012, Doyle and

² For example: ‘A bat and a ball together cost \$1.10. The bat costs \$1 more than the ball. How much does the ball cost?’ (Not \$0.10.)

Hourihan 2016), or spaced vs. massed training (Eglington and Kang 2017, Wahlheim, et al. 2011, Yan, et al. 2016, Rawson, et al. 2015). Little is known about the effects of category learning judgements on downstream cognition. For example, are people less likely to use a concept if they don't think they have learnt the category very well? People do use their CLJs when selecting items that they would like to study again. But the effect differs between individuals. Some people select for further study categories that they judged to be least well learned, others select the categories that they judge to be most well learned (Morehead, et al. 2017).

More recently my collaborators and I have examined a range of ways that thinkers appraise their concepts epistemically. We did find evidence for an effect of monitoring on control. Concepts in which people tend to report high confidence are also those that are selected for performing inductions (Thorne, et al. 2021). But we did not investigate whether inconsistency in judgements has an impact on concept appraisal, along the lines suggested by Millikan's hypothesis. We did find that the coherence of the causal structure of a category relates to the way people appraise a concept (Thorne, et al. 2022), so a link with consistency is not implausible, but not something we can determine from our data.

Taking these results in the round I would conclude that there is good evidence that beliefs and judgements are monitored for consistency but little evidence in support of Millikan's hypothesis that inconsistency in judgement makes the thinker less disposed to rely on the concepts involved or to rely the ways the concepts were applied. Millikan's claim here is entirely plausible given the effects that have been found to date, but more evidence is needed if we want to give consistency testers a central role in our theorising.

(7) Role in the Evolution of Concepts and Language

Culture changes over time. Tools, art, behavioural practices, ideas, concepts and language are all passed between people in different ways and spread over time in cultural groups. If 'evolution' just means change over time, then all of these aspects of culture evolve. A more interesting hypothesis is that culture evolves in something like the way genetically-transmitted traits do, that is, through evolution by natural selection. The suggestion that ideas, concepts and language evolve in this way is familiar from (Dawkins 1976). Dawkins named the so-transmitted cultural items 'memes' (a term that has subsequently come to mean a picture, video clip or idea that is circulated widely online through social media).

One study of the way people use and share labels to categorise ambiguous stimuli found that, although single individuals and small groups adopt a diverse range of categorisation systems, large groups converge on similar categorisation systems (Guilbeault, et al. 2021). This convergence in the way people use linguistic labels in categorisation suggests that there will also be convergence in the way individuals sort the stimuli under concepts. If concepts do evolve in this way, what is the selection mechanism? What are they selected for? Some concepts spread widely in a culture,

others surface briefly but then die out. The same is true of new linguistic terms. For the claim that there are memes to offer any insight into how and why concepts spread, we need to understand the selection mechanism or selection mechanisms at work. Only if there is selection at work in some sense is there any profit in the analogy with gene-based natural selection.

Consistency testers would be such a selection mechanism. Concepts undergo a selective process within an individual thinker. They are selected for consistency. The concepts that persist are those that can be reliably applied in a variety of different ways. Metaphysically, this means that there is an advantage to concepts that refer to 'substances' in Millikan's (broadly Aristotelian) sense: entities that have a range of predictable properties, hence support a range of inductions (Millikan 2000). Individuals, stuffs and natural kinds are paradigmatic substances. Millikan emphasises that where different individuals have concepts of the same referent they are likely to differ in how they apply those concepts and in the beliefs that attach to them (i.e. unicepts are not shared). Nevertheless, different individuals do have concepts of the same referents. As Millikan emphasises, people often acquire concepts by hearing someone else use a linguistic label ('that is Hemara'). If humans have consistency testers we should expect concepts of substances to have an advantage in the epidemiology of cultural transmission.

There are doubtless myriad influences on the way concepts spread. Factors as diverse as prestige, usefulness, moral norms and the whims of fashion are all doubtless in play. Some of these factors will have no consistent effect over time. They will show up as noise in the transmission process. Factors which have a consistent effect can do more work to explain cultural evolution. Consistency testing promises to be one such factor. As such, it could have a considerable influence on the cultural evolution of our conceptual repertoire. Whether consistency testers exist, and how they work should, then, be of considerable interest to cultural evolutionary theorists.

Millikan argues that one of the main ways in which people acquire a new concept is by learning a new word (2017, 99). That is surely plausible. A concept acquired in this way then has to face the tribunal of consistency testing. Only if it survives will it remain in the new speaker's repertoire to be passed on in turn. So consistency testing is a filter on the transmission of linguistic terms in a culture. As with concepts, there are doubtless myriad influences on the proliferation of linguistic terms: language-specific features like phonetic felicity, in addition to all those that apply to the associated concepts. However, consistency testing would be part of the picture and, since its effects are consistent, it may play an important explanatory role.

This all suggests that, if the empirical evidence stacks up in favour of the existence of consistency testers playing the cognitive role Millikan hypothesises, then their function will have an important effect on the cultural evolution of words and concepts. The function of consistency testers is to check that the thinker has a range of reliable ways of picking out things in the world, and that the things picked out are objective entities with a wide range of predictable properties. If consistency testers are at work in selecting words and concepts, then we should expect words and

concepts which pick out such entities to be at an advantage, other things being equal, in being able to proliferate in human cultures.

(8) Conclusion

Millikan's claim about consistency testing is original and important. The need to ascertain whether concepts are equivocal or redundant is a theme running through much of her work. One of her many original contributions is to argue that a mechanism that tests for consistency between judgements is, in the first instance, not a test of the truth of judgement (as previously assumed), but a way of checking whether concepts are univocal. Consistency testers also play a central role in Millikan's theory of the representational content of judgements, and thus of the concepts out of which judgements are composed. Although other aspects of how concepts are formed and used may also play a role, it is plausible that consistency testing is part of what endows concepts with determinate content.

Consistency testing in the specific form proposed by Millikan has not been directly investigated empirically. The hypothesis is however plausible in the light of existing evidence about metacognition, and the large body of evidence about expectations and error signals. It may be, however, that the phenomenon is more general than Millikan suggests – for instance, not always depending on deploying a concept of negation. If the empirical evidence does in time stack up in favour of the idea that concepts face an internal tribunal of consistency testing, or something like it, that finding will be extremely important to understanding the cultural evolution of words and concepts.

Acknowledgements

For discussion and comments on previous drafts, the author is very grateful to Asher Koriat, Joëlle Proust and Wayne Wu; audiences at the Oberlin Colloquium in Honour of Ruth Millikan, an ANU graduate 'post-talk', and the Institute of Philosophy lab meeting; and especially to Ruth Millikan, for extensive comments on a draft of the paper, and for generous discussion of these issues over many years. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681422 (MetCogCon).

References

- Ackerman, R., and V. A. Thompson. 2017. "Meta-Reasoning: Monitoring and Control of Thinking and Reasoning," *Trends in Cognitive Sciences*, **21**: 607-17.
- Astuti, R., and P. L. Harris. 2008. "Understanding Mortality and the Life of the Ancestors in Rural Madagascar," *Cogn Sci*, **32**: 713-40.
- Dawkins, R. 1976. *The Selfish Gene*. Oxford: O.U.P.

- Doyle, Mario E., and Kathleen L. Hourihan. 2016. "Metacognitive Monitoring During Category Learning: How Success Affects Future Behaviour," *Memory*, **24**: 1197-207.
- Eglington, Luke G, and Sean HK Kang. 2017. "Interleaved Presentation Benefits Science Category Learning," *Journal of Applied Research in Memory and Cognition*, **6**: 475-85.
- Fernandez Cruz, A. L., S. Arango-Munoz, and K. G. Volz. 2016. "Oops, Scratch That! Monitoring One's Own Errors During Mental Calculation," *Cognition*, **146**: 110-20.
- Friston, K. J., and K. E. Stephan. 2007. "Free-Energy and the Brain," *Synthese*, **159**: 417-58.
- Godfrey-Smith, Peter. 2013. "Signals, Icons, and Beliefs". In Ryder, Kingsbury and Williford, eds, *Millikan and Her Critics*. Oxford / Malden MA: Wiley-Blackwell, 41-58.
- Guilbeault, D., A. Baronchelli, and D. Centola. 2021. "Experimental Evidence for Scale-Induced Category Convergence across Populations," *Nat Commun*, **12**: 327.
- Hamm, Jeff P, Blake W Johnson, and Ian J Kirk. 2002. "Comparison of the N300 and N400 Erps to Picture Stimuli in Congruent and Incongruent Contexts," *Clinical neurophysiology*, **113**: 1339-50.
- Hartwig, Marissa K, and John Dunlosky. 2017. "Category Learning Judgments in the Classroom: Can Students Judge How Well They Know Course Topics?," *Contemporary Educational Psychology*, **49**: 80-90.
- Jacoby, Larry L., Christopher N. Wahlheim, and Jennifer H. Coane. 2010. "Test-Enhanced Learning of Natural Concepts: Effects on Recognition Memory, Classification, and Metacognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**: 1441.
- Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, and Anna Potapenko. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature*: 1-11.
- Koriat, Asher. 1997. "Monitoring One's Own Knowledge During Study: A Cue-Utilization Approach to Judgments of Learning," *Journal of Experimental psychology: General*, **126**: 349.
- . 2011. "Subjective Confidence in Perceptual Judgments: A Test of the Self-Consistency Model," *Journal of Experimental psychology: General*, **140**: 117-39.
- . 2012a. "The Self-Consistency Model of Subjective Confidence.," *Psychological Review*, **119**: 80-113.
- . 2012b. "The Subjective Confidence in One's Knowledge and Judgments: Some Metatheoretical Considerations". In Beran, Brandl, Perner and Proust, eds, *The Foundations of Metacognition*. Oxford / New York: OUP, 213-33.
- . 2012c. "The Relationships between Monitoring, Regulation and Performance," *Learning and Instruction*, **22**: 296-98.
- Kornell, Nate, and Robert A Bjork. 2008. "Learning Concepts and Categories: Is Spacing the "Enemy of Induction"?," *Psychological Science*, **19**: 585-92.

- Kornell, Nate, Alan D Castel, Teal S Eich, and Robert A Bjork. 2010. "Spacing as the Friend of Both Memory and Induction in Young and Older Adults," *Psychology and aging*, **25**: 498.
- Kutas, Marta, and Kara D Federmeier. 2011. "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (Erp)," *Annual review of psychology*, **62**: 621-47.
- Legare, Cristine H., and Andrew Shtulman. 2018. "Explanatory Pluralism across Cultures and Development". In Proust and Fortier, eds, *Metacognitive Diversity: An Interdisciplinary Approach*. Oxford: OUP.
- Martinez, Manolo. 2013. "Teleosemantics and Productivity," *Philosophical Psychology*, **26**: 47-68.
- Millikan, Ruth Garrett. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
- . 1989. "Biosemantics," *Journal of Philosophy*, **86**: 281-97.
- . 2000. *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
- . 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford / New York: OUP.
- . 2022. "Self-Signs and Intensional Contexts," *Mind & Language*: 1-19.
- Morehead, Kayla, John Dunlosky, and Nathaniel L Foster. 2017. "Do People Use Category-Learning Judgments to Regulate Their Learning of Natural Categories?," *Memory & Cognition*, **45**: 1253-69.
- Nelson, Thomas O., and Louis Narens. 1990. "Metamemory: A Theoretical Framework and New Findings". In Bowerman, ed, *The Psychology of Learning and Motivation*. Cambridge, MA: Academic Press, Inc., 125-73.
- Polich, John. 2007. "Updating P300: An Integrative Theory of P3a and P3b," *Clinical neurophysiology*, **118**: 2128-48.
- Proust, J. 2008. "Epistemic Agency and Metacognition: An Externalist View," *Proceedings of the Aristotelian Society*, **108**: 241-68.
- . 2012a. "The Norms of Acceptance," *Philosophical Issues*, **22**: 316-33.
- . 2012b. "Metacognition and Mindreading: One or Two Functions?". In Beran, Brandl, Perner and Proust, eds, *Foundations of Metacognition*. Oxford: OUP, 234-51.
- . 2013a. *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford University Press.
- . 2013b. "Mental Acts as Natural Kinds". In Vierkant, ed, *Decomposing the Will*. Oxford / New York: OUP, 262-82.
- Rawson, Katherine A., Ruthann C. Thomas, and Larry L. Jacoby. 2015. "The Power of Examples: Illustrative Examples Enhance Conceptual Learning of Declarative Concepts," *Educ Psychol Rev*, **27**: 483–504.
- Rupert, Robert D. 1999. "Mental Representations and Millikan's Theory of Intentional Content: Does Biology Chase Causality?," *The Southern journal of philosophy*, **37**: 113-40.
- Sperber, Dan, Fabrice Clement, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. "Epistemic Vigilance," *Mind & Language*, **25**: 359-93.
- Thompson, Valerie A., and Stephen C. Johnson. 2014. "Conflict, Metacognition, and Analytic Thinking," *Thinking & Reasoning*, **20**: 215-44.

- Thorne, S. R., J. Quilty-Dunn, J. Smortchkova, N. Shea, and J. A. Hampton. 2021. "Concept Appraisal," *Cogn Sci*, **45**: e12978.
- Thorne, S. R., J. Smortchkova, J. Quilty-Dunn, N. Shea, and J. A. Hampton. 2022. "Is Concept Appraisal Modulated by Procedural or Declarative Manipulations?," *Front Psychol*, **13**: 774629.
- Wahlheim, Christopher N, John Dunlosky, and Larry L Jacoby. 2011. "Spacing Enhances the Learning of Natural Concepts: An Investigation of Mechanisms, Metacognition, and Aging," *Memory & Cognition*, **39**: 750-63.
- Wahlheim, Christopher N, Bridgid Finn, and Larry L Jacoby. 2012. "Metacognitive Judgments of Repetition and Variability Effects in Natural Concept Learning: Evidence for Variability Neglect," *Memory and Cognition*, **40**: 703-16.
- Yan, Veronica X, Elizabeth Ligon Bjork, and Robert A Bjork. 2016. "On the Difficulty of Mending Metacognitive Illusions: A Priori Theories, Fluency Effects, and Misattributions of the Interleaving Benefit," *Journal of Experimental psychology: General*, **145**: 918.