THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
Vol. 69, No. 11, 2166–2188, http://dx.doi.org/10.1080/17470218.2015.1106565

Event-related potentials reveal the relations between feature representations at different levels of abstraction

Samuel D. Hannah[1,2], Judith M. Shedden[1], Lee R. Brooks[1], and John G. Grundy[1]

[1]Department of Brain, Behaviour & Neuroscience, McMaster University, Hamilton, ON, Canada
[2]Department of Psychology, University of Saskatchewan, Saskatoon, SK, Canada

(Received 26 January 2015; accepted 14 July 2015; first published online 9 November 2015)

In this paper, we use behavioural methods and event-related potentials (ERPs) to explore the relations between informational and instantiated features, as well as the relation between feature abstraction and rule type. Participants are trained to categorize two species of fictitious animals and then identify perceptually novel exemplars. Critically, two groups are given a perfectly predictive counting rule that, according to Hannah and Brooks (2009. Featuring familiarity: How a familiar feature instantiation influences categorization. *Canadian Journal of Experimental Psychology/Revue* Canadienne de Psychologie Expérimentale, *63*, 263–275. Retrieved from http://doi.org/10.1037/a0017919), should orient them to using abstract informational features when categorizing the novel transfer items. A third group is taught a feature list rule, which should orient them to using detailed instantiated features. One counting-rule group were taught their rule before any exposure to the actual stimuli, and the other immediately after training, having learned the instantiations first. The feature-list group were also taught their rule after training. The ERP results suggest that at test, the two counting-rule groups processed items differently, despite their identical rule. This not only supports the distinction that informational and instantiated features are qualitatively different feature representations, but also implies that rules can readily operate over concrete inputs, in contradiction to traditional approaches that assume that rules necessarily act on abstract inputs.

Keywords: Abstraction; Amodal thought; Feature processing; Instantiated features; Rules.

Researchers investigating categorization and concept formation have traditionally treated an exemplar's features as comprising two classes of information. Features could be category relevant, in which case they helped define an item as a member of a category, or features could be idiosyncratic, in which case they helped define an item as an individual. Category-relevant features have typically been treated as generic features, while individuating information is equated with the specific perceptual form of features (e.g., Medin, Dewey, & Murphy, 1983; Regehr & Brooks, 1993). By "generic feature" we mean the feature content that holds constant not only across different exemplars,

Correspondence should be addressed to Samuel Hannah, Department of Psychology, University of Saskatchewan, Saskatoon, SK, Canada and Samuel Hannah is now in private industry. E-mail: hannah.sam@gmail.com

Samuel Hannah is now at the University of Saskatchewan. John Grundy is now at York University. Lee Brooks is deceased.

Lee Brooks passed away after an earlier draft of this paper was written; this paper owes much to Lee for its merits. The paper's shortcomings, however, belong only to the first two authors, who also owe much to Lee.

This research was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to J.M.S. and to L.R.B.

but even across different categories. Cats, dogs, and monkeys all have paws—there is a set of properties common across all these end-limb structures such that it is meaningful to call them all "paw" (Figure 1); "paw" defined this way is a generic feature. Brooks and Hannah (2006), however, argued that the specific perceptual forms of features found in natural categories also carry category-relevant information because such forms are typically category specific. The paws of individual cats look quite different in terms of the details of their appearance—what we call "feature form"—from the paws of dogs or monkeys, and the paws of individual cats look similar in detail to one another, especially when compared to those of other animals. That feature forms are category specific makes them sufficient for categorization, and this give feature forms great potency: The briefest flash of a cat's paw lunging around a corner is sufficient to tell you what creature lies in wait for your toes.

Concept formation and category use are served, argued Brooks and Hannah (2006), not only by mental representations capturing the generic features ("informational features"), but also by mental representations capturing the feature form ("instantiated features"). We may loosely think of instantiated features as feature instances and informational features as something like feature prototypes.

Consider Figure 1: At the top are a number of paws, all from the same category, that of "cat". We could represent the fact that cats have paws by a mental representation of one of the specific feature forms in the top row, say, paw in the leftmost picture. Such a mental representation would be an instantiated feature. However, along the bottom are a number of paws from different categories, each manifesting a strikingly different specific form. Despite their differences, however, there is some paw information present in each of these different paw types. We could represent the fact that cats have paws with a mental representation of this generic paw information. Such a mental representation would be an informational feature.

If category-relevant feature information is found across multiple levels of abstraction, as suggested by Brooks and Hannah (2006), then several questions become important. For example, we need to ask how the contents of informational features relate to the contents of instantiated features, and thus

Figure 1. *A feature, such as paw, has a specific form that is typically category specific in natural categories (top row). However, most natural features also embody more generic information that is only apparent when comparing across categories (bottom row). We can represent the feature of a category in terms of either its specific feature form or its generic content.*

how informational features may be generated from the specific forms encountered. One possibility is that informational features are generated automatically as a sparser, or fuzzier, version of some corresponding instantiated feature. That is, informational features may be broadly analogous to Brainerd and Reyna's (1990a, 1990b, 1996; Reyna & Brainerd, 1995) gist traces of fuzzy-trace theory, while instantiated features would be similar to verbatim traces. The two classes of feature representation would differ only quantitatively.

Alternatively, informational features could emerge out of a more deliberate comparison across features that have been aligned across a set of relevant dimensions so as to reveal commonalities, differences, and relations across the set. This would produce informational features that capture qualitatively different information from that carried by instantiated features. Structural alignment as a means of determining similarities has been an influential idea emerging out of the work of Gentner and Markman (e.g., Gentner & Markman, 1997; Markman & Gentner, 1993, 1997), and this has been extended to account for the emergence of some types of abstractions. Gentner and Medina (1998) argued that generic constructs such as rules could emerge from such a process of aligned comparison, and Yamauchi (2009) has more recently argued that abstract feature representations require aligned comparisons as well. If informational features emerge from some sort of deliberate comparison, then the two classes of feature representation should differ qualitatively.

In this paper, we use behavioural and electrophysiological methods to clarify the relation between informational and instantiated features. In order to see differential responding to informational and instantiated features, we need transfer materials that allow decision making to be influenced by either instantiated or informational features. That is, we need transfer items that have one or more features similar or identical to features learned in training, as well as transfer items without any such perceptually familiar features. To do this, we modify the training and transfer materials used by Brooks and Hannah (2006; Hannah & Brooks, 2006, 2009). Event-related potential (ERP) responses may prove even more sensitive to these processing distinctions than behavioural measures. People reliant on instantiated features to make decisions may show an effect of transfer item in their ERP activity after initial perceptual processing but before response programming, while those reliant on informational features should show no such intermediate-to-late effect.

If informational and instantiated features differ only in the amount of detail they capture, then we would expect to see similar behavioural and ERP responses across transfer items regardless of whether people are reliant on informational or on instantiated features to make their categorization decisions. However, if informational features are qualitatively different from instantiated features, then we would expect that people reliant on informational features should show different ERP responses compared with those reliant on instantiated features. As this is exploratory work, it is unclear what form these ERP differences may take; to guard against biases introduced by a visual examination of electrode activity, we selected electrodes for analysis using an objective technique called partial least-square analysis (PLS; McIntosh, Bookstein, Haxby, & Grady, 1996).

In order to carry this analysis out, we need to have a means of identifying or generating participants reliant on informational features and participants reliant on instantiated features when making categorization decisions. Brooks and Hannah (2006; Hannah & Brooks, 2006, 2009) have argued that the different feature types have different relations with types of rules. Thus, by giving people different types of rules to use, we hoped to create participants reliant upon different types of feature representations. This manipulation allowed us to explore the relations between rules and feature representations as a secondary topic.

Types of features are related to types of rules

Sloman (1996, p. 5) argued that, "Rules are abstractions that apply to any and all statements that have a certain well-specified, symbolic structure. Most important, they have both a logical structure and a set of variables." Thus, rules are (a) abstractions (b) operating on abstractions ("symbolic structure") and (c) have a logical organization that determines a specific integration of inputs. The inputs, being variables, vary in their specific instantiation, and therefore what rules act on is the abstract content of these variable inputs: Rules *must* operate on abstract inputs. In Brooks and Hannah's (2006) terminology, rules must operate on informational features.

Applied to the categorization of physical objects, this view of rules implies a process in which perceptual inputs are transformed into abstract symbols and integrated by a logical structure, leading to an inevitable answer. For example, the rule "If an animal has two of rounded head, rounded torso, striped coat or two legs, then it is a bleeb", will always produce the categorization "bleeb" whenever a creature with at least two of those four features is encountered, and will never produce the category "bleeb" if only one of the listed features is encountered in an animal. Brooks and Hannah (2006) called this a "strong rule", because it specifies how features are combined—in this case, by counting features—to lead to a decision. If some of the bleeb-type features are encountered in another creature, our counting rule will distinguish the pseudo-bleeb from the true bleeb. This is the result of the "logical structure and a set of variables" that Sloman (1996) emphasizes.

When asked to define natural categories, however, people frequently provide simple feature lists (Rosch & Mervis, 1975)—for example: "A bleeb usually has a rounded head, rounded torso, striped coat or two legs." If we assume that there is some implicit logical structure to such feature lists, perhaps an unstated counting rule, then such definitions are unproblematically rules. Brooks and Hannah (2006), however, argued that such feature lists are just what they appear to be: a list of features without any implicit logical structure for integrating the features into a conclusion. These work for a person generating the list because the features named are the category-specific, and hence sufficient, feature forms used to make decisions. Feature lists, argued Brooks and Hannah, are pointers to the instantiated features a decision-maker has relied on. Nonetheless, they perform rule-like functions, most especially by providing a scaffolding for learning by directing attention to key regions. If you are told a bleeb has a rounded head, you are unlikely to know precisely what a bleeb's head look likes. You are likely, however, to know enough about rounded heads to have some broad parameters as to what a bleeb head should look like, and to know where to look to confirm its actual shape. Thus, Brooks and Hannah called feature lists "weak rules".

For the rest of this paper, we refer to strong rules as "counting rules", as these are the only type of strong rules we explore. Similarly, we will refer to weak rules as "feature lists", as these are the only type of weak rules we explore. However, readers should keep in mind that other forms of strong and weak rules are possible.

Hannah and Brooks (2009) explored the relations between feature and rule types; they trained people to classify four species of fictitious animals and then transferred them to classifying novel exemplars. They found that participants who provided counting-rules after the transfer test behaved largely as if they were adding up the features they detected and erring by missing features. In contrast, participants who provided feature lists seemed to weight features by their "goodness", or recognizability. Such a feature-goodness heuristic would be easier to apply when working with instantiated features rather than informational features because informational features are sparse and invariant—any rounded-head informational feature is identical to, and thus just as good as, any other rounded-head informational feature.

Further, Hannah and Brooks (2009) gave a group of yoked participants the counting rules or feature lists generated by other participants, but denied these yoked participants exposure to the training stimuli prior to the classification task. Thus, the only source for learning about the categories came through the provided rules, and not through any perceptual experience. Importantly, the yoked participants who were given feature lists behaved identically to the counting-rule producers and showed no evidence of the feature-weighting typical of the people who had generated the provided feature lists. It appeared, therefore, as if a feature-weighting strategy requires a stock of instantiated features. The yoked feature-list classifiers, operating without such a stock, seemed to have interpreted the provided feature lists as implicit counting rules.

Hannah and Brooks (2009) suggested, therefore, that counting rules work on informational features, as Sloman (1996) suggested, because informational features tend to overlap across categories (e.g., four legs is a possible feature across multiple categories), creating confusions that necessitate an explicit rule for integrating features. That counting rules reflect how features are distributed implies a looser relation between rules and feature representations than the description given by Sloman, for whom counting rules are *necessarily* linked to informational features. Instead, Hannah and Brooks are suggesting that counting rules are only contingently related to informational features. In contrast, the category-specific nature of instantiated features does not typically result in feature overlap across categories and therefore does not demand a logical structure to resolve conflicts as no conflicts exist. Instead, the sufficiency of instantiated features means that if a bleeb-like rounded head is spotted, then the creature is certainly a bleeb. In this formulation, then, counting rules could operate over any type of feature representation, regardless of its level of abstraction.

Experiment overview

We based our experimental strategy around this posited relation between counting rules and informational features, and feature lists and instantiated features. Like Brooks and Hannah (2006), we trained participants to categorize members of two imaginary animals (bleebs and ramuses) characterized by category-specific feature forms. Participants then engaged in a transfer task, categorizing items that were either entirely perceptually novel (all-novel items), or contained a single feature form encountered in training exemplars (interfering- or facilitating-familiarity items). Unlike Hannah and Brooks (2009), we gave participants rules to use. One group of participants was given a feature-list rule immediately after training and before test (list-after group); two other groups were given a counting rule, with one group receiving it before training (count-before group), and one immediately after training (count-after group). We tried to ensure use of the assigned rule by requiring perfect recitation of it prior to testing and confirmed use via debriefing after testing. The exact wording of the rules for the bleeb category is given in the Method section, and the debriefing questions are provided in Appendix A.

Hypotheses

Behaviourally, we can expect that both counting-rule groups should be slower but more accurate than the list-after group. The two counting-rule groups have a perfect rule, but the list-after group does not, which should yield higher accuracy for the two counting-rule groups. The counting rule, however, requires that at least two features be inspected, and also counted, while the list-after group is free to check a single feature if they wish, and no counting is required, which should yield slower responses for the two counting-rule groups.

Based on the relations between rule types and feature types described above, we expected that the list-after group should be reliant on instantiated features when categorizing the transfer items. The count-before group should be reliant on informational features, as their rule provided them with set of generic feature descriptions to guide their learning. If informational and instantiated features are qualitatively different, then when we manipulate the familiarity of transfer items' features, we should see these two groups show qualitatively different ERP responses to the manipulation. With sufficient power, we should see the list-after participants show differential ERP responses to transfer items with familiar and unfamiliar feature forms, while count-before participants show no such differential responding; at a minimum the list-after group should show evidence of greater processing as they engage more detailed feature representations (instantiated features).

The contrast between the count-before and count-after group is also interesting given that they share the same decision rule. If the relationship between counting rules and informational features is one of necessity, as implied by the traditional rule description (e.g., Sloman, 1996), then both groups should be reliant on informational features and look identical on both behavioural and ERP measures. In contrast, if count-after participants are applying their rule to instantiated features acquired in training, then count-after participants should look more like list-after participants on ERP measures. Of course, this expectation also presumes that instantiated and informational features are qualitatively different, and therefore the contrast between count-before and count-after groups is also a test of that more general hypothesis.

We did not include a list-before group for two reasons, even though this would allow us to investigate the relation between instantiated features and weak rules. First, while this an interesting topic, it is secondary to the issues of the relations between informational and instantiated features and the nature of the inputs a strong rule can act on. Second, it is difficult to give people an explicit instruction on features prior to training and prevent a strong rule from being discovered and used (Hannah & Brooks, 2009). Thus a majority of participants given a feature-list rule prior to start of training would be likely to discover a strong rule by the end of training and may then continue to use it on at least some transfer trials.

ERP-motivated modifications

We modified Brooks and Hannah's (2006, Experiment 1) transfer task to make it amenable to ERP. Brooks and Hannah trained participants to classify artificial categories very similar to those used in our task. Unlike Brooks and Hannah, we manipulated the familiarity of transfer item feature forms within subjects. ERPs require a large number of observations to produce reliable ERP averages; therefore, we also created many versions of each transfer item. To increase the range of responding and the chance of finding effects, we also included a condition in which the single familiar feature was consistent with the correct categorization, whereas Brooks and Hannah used only interfering-familiarity and all-novel conditions. Thus, people saw 12 perceptually different versions of what at an informational level was the same transfer item.

EXPERIMENTAL STUDY

Method

Participants

We wanted to ensure that participants were trained on the material and used the rule assigned to them; this is a standard concern on concept learning experiments, made more urgent by the low signal-to-noise ratio inherent in ERP data. We therefore established more stringent inclusion conditions than are in place for most such experiments. Importantly, these requirements were established at the start of the experiment. Rejections were based purely on the following behavioural grounds, and prior to inspecting the ERP data.

We followed convention in requiring that participants meet a learning criterion (e.g., Markman & Maddox, 2003; Sweller & Hayes, 2010). We used Brooks and Hannah's (2006) criterion of over 80% accuracy on the final round of training; this level reflects a level of performance greater than can be achieved by reliance on a single feature. We also required that participants' transfer accuracy on the all-novel transfer items for both categories be above chance. To further ensure that all participants analysed had understood their rules and learned the categories, we analysed accuracy for outliers, dropping participants whose accuracy was more than three standard deviations below their group means across all three test-item conditions. We used a recursive outlier detection procedure applied to the accuracy measures to identify outliers, selecting candidate outliers using an initial liberal criterion of two standard deviations below the group mean across all three test-item conditions. We recalculated the means and standard deviations after removing candidate outliers, which were then tested against the revised means using a conservative three-standard-deviation rule and the revised means and standard deviations. A recursive procedure guards against using thresholds that have themselves been contaminated by outliers.[1] Finally, we debriefed all participants to ensure that they followed the rule assigned to them, dropping those who did not (debriefing protocol in Appendix A).

Sixty-three McMaster undergraduate students initially participated in exchange for credit in first- or second-year psychology courses. We dropped two participants in the list-after group for failing to reach learning criterion. Six participants were dropped in the count-after group, two for failing to use their counting rule at test, and four for being outliers on accuracy. We dropped seven participants from the count-before group, one for failing to reach training criterion, and six for being outliers on accuracy. No participants failed the test criterion. This left a total of 48 participants, with 16 participants randomly assigned to list-after, count-after, and count-before groups. The rather high level of outliers may reflect a combination of the difficulty of using the counting rule under speeded conditions combined with the distraction of an ERP set up. Unlike participants in standard categorization-learning experiments, our participants wore an electrode cap with 128 gel-filled electrodes, plus four additional face electrodes; a chin bar restricted motion, and participants were asked to control blinking. It is not unexpected to observe greater distraction and lower performance than normal, especially when coupled with a fairly challenging learning task.

Stimuli, apparatus, and procedure

The experiment was conducted on a 2.4-GHz Pentium 4 computer, running the Windows 2000 operating system. Stimuli were displayed on a Samsung SyncMaster 17″ colour monitor, with a refresh rate of 75 Hz and a screen resolution of 1024 × 768 pixels. The screen was set 80 cm from the participant; a chin rest fixed viewing distance. Stimulus presentation, timing, and response collection was controlled by Presentation software (http://nbs.neuro-bs.com/).

Training and transfer stimuli. We used training materials typical of those used in Brooks and Hannah's previous experiments (Brooks & Hannah, 2006; Hannah & Brooks, 2006, 2009), and they consisted of line drawings of imaginary animals composed of two species defined by four dimensions taking on binary informational features. Each training category consisted of five exemplars: a prototype with all four features typical of the category (shown at the top row in Figure 2), and four nonprototype items that differ from the prototype by a single feature.

Our training materials had a family-resemblance organization in terms of their generic features, as illustrated in Table 1. By "family resemblance", we mean that the categories lacked any single feature that is common to all members of a category, or that are exclusive to a single category (Rosch & Mervis, 1975). As Table 1 shows, for example, most bleebs had at least three of rounded head, rounded torso, striped coat, and two legs. However, all but one bleeb—the category prototype, shown in the top row of Table 1—took on the typical value for ramuses on one feature such that ramus features overlapped with bleebs. For example, one bleeb has four legs, another has an angular head, and so on. Bleeb features similarly show up among ramuses: One ramus has two legs, another has a rounded head, and so on. These overlapping features we call "lure features". In contrast, features that are typical of the membership category we call "characteristic features".

Note that all training feature forms were category specific such that lure features only existed as generic features. For example, the bleeb in the second row of Figure 2 has an angular head, which is a lure feature because angular head is typical of ramuses. The particular form of its angular head, however, looks nothing like the angular heads that appear among ramuses, which are identical with one another.

Transfer items consisted of three types of new stimuli distinguished by the perceptual familiarity

[1]As Van Selst and Jolicoeur (1994) showed, a recursive threshold applied to reaction time data (or any ex-Gaussian distributed data) yields worse performance than simple threshold procedures. However, we are only applying a recursive method on proportion correct, which does not follow an ex-Gaussian distribution.

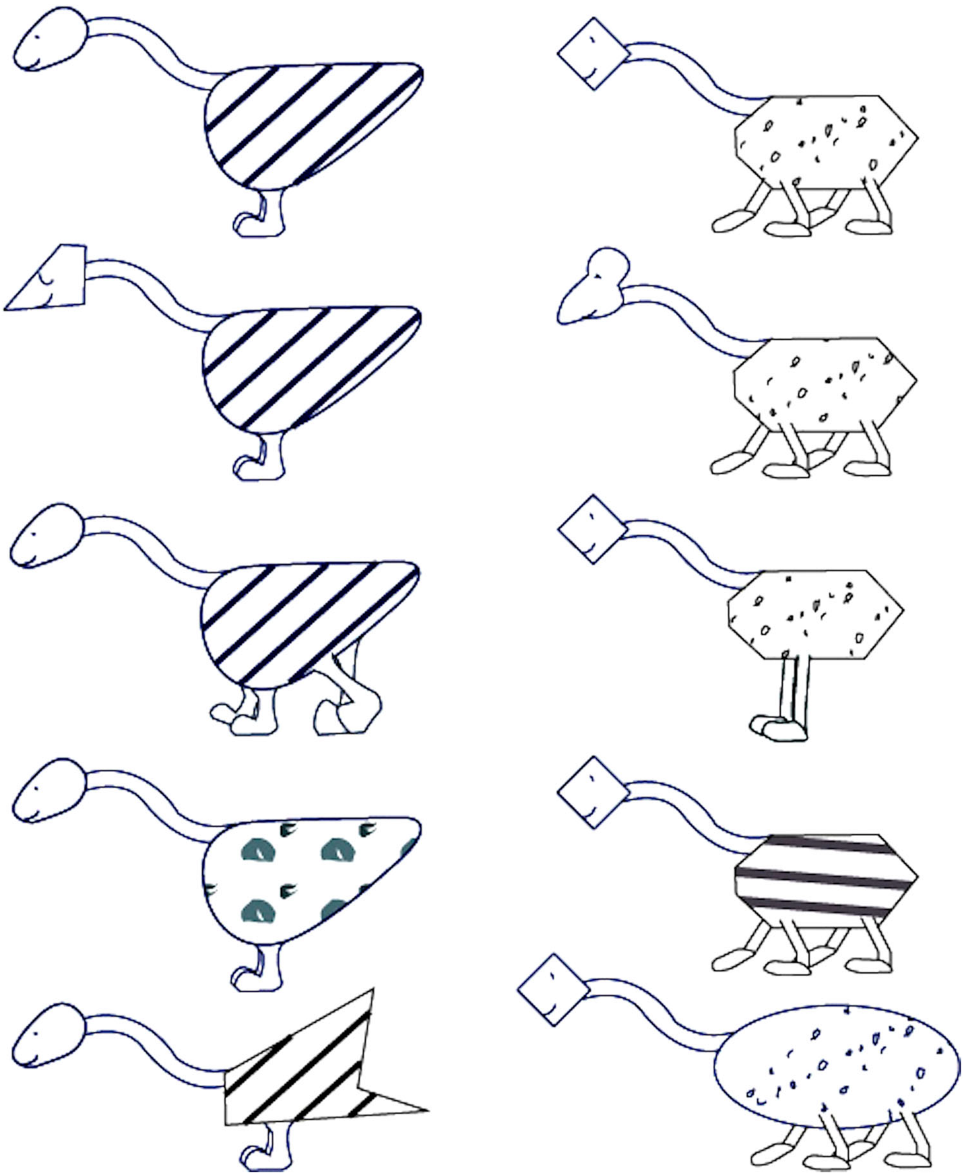

Figure 2. *Training materials. Bleebs are shown in the left column and ramuses in the right. Category prototypes are shown in the first row. Nonprototype items (one-away items) deviated from their prototypes by taking on the characteristic generic value of the rival category along one dimension (row 2: angular/round head; row 3: two/four legs; row 4: spotted/striped pattern; row 5: angular/round torso).*

of their characteristic and lure features (bottom row, Figure 3). The all-novel transfer items consisted solely of perceptually novel versions of the generic features of the nonprototype training items. The all-novel item in Figure 3, for example, is a novel manifestation of the bleeb training item shown in the row above it, and both correspond to Item 2 in Table 1. For the facilitating-familiarity items, one of the characteristic features took on the category-specific form encountered in

Table 1. *Informational structure of the training set*

		Features			
Category	*Items*	*Head*	*Torso*	*Pattern*	*Leg no.*
Bleeb	1 (prototype)	1	1	1	1
	2	0	1	1	1
	3	1	0	1	1
	4	1	1	0	1
	5	1	1	1	0
Ramus	6 (prototype)	0	0	0	0
	7	1	0	0	0
	8	0	1	0	0
	9	0	0	1	0
	10	0	0	0	1

Note: Head: 1 = rounded, 0 = angular; torso: 1 = rounded, 0 = angular; pattern: 1 = stripes, 0 = dots; number of legs: 1 = two legs; 0 = four legs.

training; the facilitating-familiarity item in Figure 3, for example, is a bleeb with the typical bleeb torso seen in training bleebs. For the interfering-familiarity items, the lure feature took on the category-specific feature form encountered in the rival training category; the interfering-familiarity item in Figure 3, for example, has the typical ramus head seen in training, even though the item is a bleeb.

We created 12 versions of the eight nonprototype training items for each of the three transfer-item conditions. For each of the transfer items shown in Figure 3, for example, there were another 11 variants, each sharing the same generic features. This yielded a total of 3 (transfer-item conditions) × 8 (one-away items) × 12 (transfer versions), or 288, transfer items.

Electrophysiology. The ActiveTwo Biosemi electrode system was used to record continuous electroencephalographic (EEG) activity from 128 silver/silver chloride (Ag/AgCl) scalp electrodes, a common mode sense (CMS) active electrode, and a driven right leg (DRL) passive electrode (www.biosemi.com/faq/cms&drl.htm). Four additional electrodes were placed at the outer

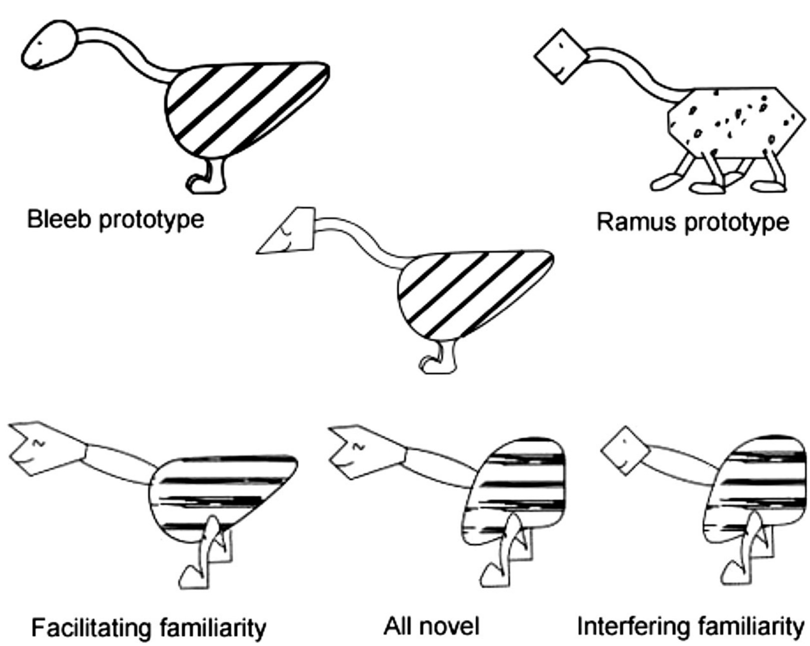

Figure 3. *Top row: bleeb (left) and ramus (right) prototypes. Middle row: nonprototype bleeb training item. Bottom row: Transfer equivalents of the nonprototype bleeb training item.*

canthi and just below each eye for recording of horizontal and vertical eye movements. The continuous signal was acquired with an open passband from DC to 150 Hz and digitized at 512 Hz. The signal was bandpass filtered offline at 0.3 to 30 Hz and re-referenced to averaged mastoid reference electrodes. Offline processing of the EEG signal was carried out with EEProbe software (ANT; www.ant-neuro.com), which was used for artefact rejection, segmenting, averaging, and eye-blink correction. A 100-ms prestimulus baseline was used for ERP averaging; only correct trials were included in the averages.

Procedure

Participants were told to categorize items by pressing an up-arrow or a down-arrow key; half of the participants used the up-arrow key to make a "bleeb" response, and the other half used the up-arrow key to make a "ramus" response. Participants used their preferred hand and fingers. Participants were instructed to respond as quickly as possible while maintaining accuracy through both training and transfer phases. Trials for both training and transfer phases began with the appearance of a fixation cross, which remained on screen until the participant hit the space bar. Items appeared centred on screen upon the participant hitting the space bar and remained until the participant made a response; participants were required to refrain from blinking during stimulus display. A schematic of the experimental procedure is given in Figure 4.

In the training phase, participants were presented with 100 training trials in three rounds, following the supported induction procedures used in Brooks and Hannah (2006). The three rounds consisted of labelled pairs (Round 1; 10 trials), unlabelled pairs (Round 2; 10 trials), and unlabelled single items (Round 3; 80 trials). Each item was presented twice in each of the two paired-item rounds. The first pairing involved items matched on lure features (e.g., rounded head ramus and angular head bleeb); the second pairing was randomly determined. Each item was presented eight times in the single-item round. We randomly ordered items in the single-item round, subject to the constraint that the first two items presented be the prototypes.

For both rounds of paired items, a prompt appeared at the top centre of the screen to indicate whether the participant was to identify the left or right pair member. For the labelled pairs in the first round, the prompt appeared five seconds after stimulus display to give participants time to study both items. In the second round, the prompt appeared simultaneous with the pair of items. Incorrect responses were signalled by a low-frequency tone, and correct responses by a high-frequency tone.

Members of the count-before group began training by rehearsing counting-based categorization rules based on the structure of the categories described in Table 1 (e.g., "An item is a bleeb if it has at least two of the following four features: four legs, rounded heads, stripes, and rounded torsos"). Members of the count-after group practised the same rule immediately after training. Members of the list-after group recited a list of four characteristic features for each category (e.g., "Bleebs usually have four legs, rounded heads, stripes, and rounded torsos") at the end of training. All participants rehearsed their rules until they could recite them perfectly. Thus, at a minimum, all participants began the transfer test knowing the relevant features for each category.

Transfer and posttransfer. All participants were told that they would see new variants of the training items and that no feedback would be given. The experimenter instructed all participants to use their assigned rule to identify transfer items and to respond as quickly and as accurately as possible. We randomly ordered transfer trials at the start of each participant's session. After completion of the test phase, participants were debriefed and were queried as to how they made their decision.

Results

We analysed mean accuracy and mean RT for correct responses in separate 3 (rule group: list-after, count-after, count-before) $\times$ 3 (transfer item: all novel, facilitating familiarity, interfering

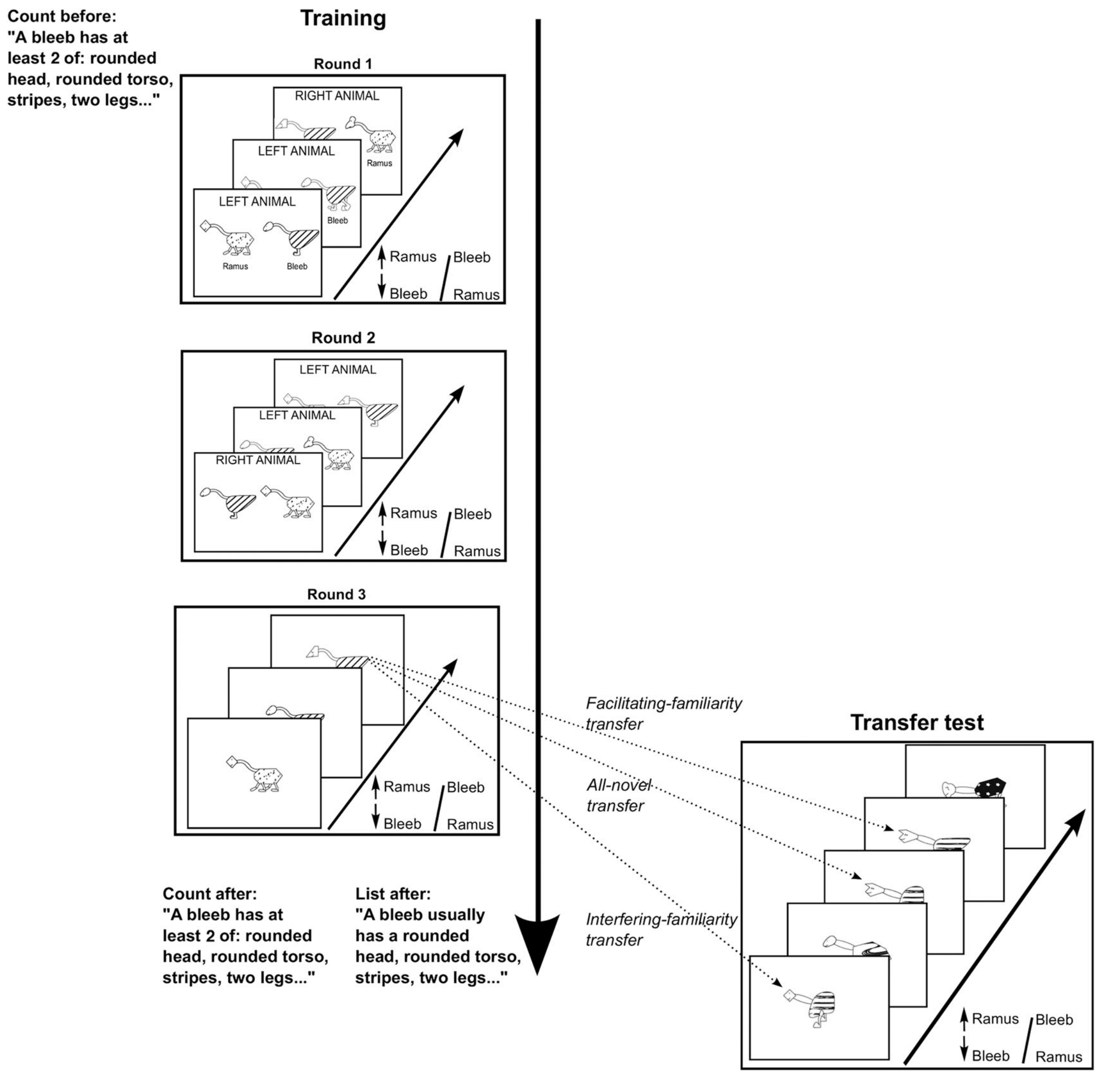

Figure 4. *Experiment schematic. Dotted arrows point to the transfer items illustrated in Figure 3. Participants categorized training and transfer items by pressing the up- and down-arrow keys.*

familiarity) × 2 (species: bleeb, ramus) mixed-design analyses of variance (ANOVAs). For both analyses, rule group was a between-subjects factor, and transfer item and species were within-subject factors. A larger analysis with response mapping as a between-subjects factor found no effect of response mapping for accuracy or reaction time (RT). Although the effect of species did not provide any meaningful results, results involving this factor are reported for completeness in Appendix B.

Accuracy

As the top panel of Figure 5 shows, the list-after group's accuracy falls below that of the two counting rule groups, which are similar to each other. For the list-after group, accuracy is highest for facilitating-familiarity items, and lowest for interfering-familiarity items, with all-novel items falling intermediate. For the two counting-rule groups, however, accuracy appears flat across transfer items. The ANOVA largely confirms these impressions, with rule group, $F(2, 45) = 70.99$, $MSE = .046$, $p < .001$, $\eta_p^2 = .76$, transfer item, $F(2, 90) = 38.90$, $MSE = .003$, $p < .001$, $\eta_p^2 = .46$, and their interaction, $F(4, 90) = 19.53$, $MSE = .003$, $p < .001$, $\eta_p^2 = .46$, achieving reliability.

To clarify the Rule Group × Transfer Item interaction, we analysed the accuracy data by conducting simple-effect analyses split on rule group. Given the post hoc nature of these analyses, we applied a Bonferroni correction to α by dividing .05 by the six possible post hoc parsings of the interaction's main effects (three transfer item parsings and three rule group parsings), $\alpha_{adjusted} = .0083$. Only the list-after group showed a reliable main effect of transfer item, $F(2, 30) = 74.88$, $MSE = .003$, $p < .001$, $\eta_p^2 = .68$. No other effects were reliable. A familiar feature appears to influence list-after participants' decisions, but not those of the counting rule participants.

It should be noted that although list-after group's accuracy (59.4%, $SE = 3.4\%$) on the all-novel was lower than that on the all-novel items in Brooks and Hannah's (2006) experiment, performance was above chance, with a 95% confidence interval[2] ranging from .522–.666. Further, all participants scored above 50% correct in classifying all-novel items for both categories; thus, not only did the list-after group as a whole perform above chance on the all-novel items, so did each participant within the group. List-after participants are able to reliably classify items in the absence of any specific similarity to guide their performance. Their knowledge is not simply restricted to a small number of instantiated features, but must also include some informational feature representations as well.

Reaction times

As illustrated in the bottom panel of Figure 5, list-after participants correctly categorized stimuli more quickly ($M = 1297$ ms, $SE = 120$) than both count-before ($M = 3116$ ms, $SE = 256$) and count-after participants ($M = 2998$ ms, $SE = 297$). This main effect of rule group proved reliable: $F(2, 45) = 20.46$, $MSE = 4{,}862{,}343$, $p < .001$, $\eta_p^2 = .48$. The main effect of transfer item was also reliable, $F(2, 90) = 5.97$, $MSE = 75{,}108$, $p = .004$, $\eta_p^2 = .12$, driven by a small advantage for facilitating-familiarity items over all-novel items, Bonferroni-corrected paired t-test ($\alpha_{adjusted} = .017$), $t(47) = 3.86$, $SE = 35.2$, $p < .001$, Hedges's $g = 0.56$.

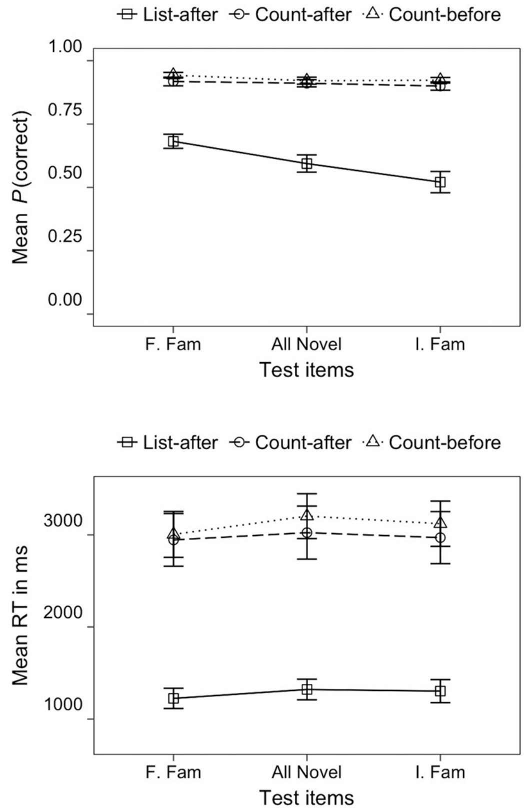

Figure 5. *Top: Mean proportion correct categorizations across transfer items by rule group. Bottom: Mean correct reaction times (RTs) across transfer items by rule group. "F. Fam" = facilitating familiarity; "I. Fam" = interfering familiarity. Error bars = 1 standard error.*

Overall, the behavioural evidence is suggestive of instantiated and informational features engaging

[2]Interval based on a critical $t(15) = 2.13$.

processing in qualitatively different ways. The feature familiarity influenced the speed by which all participants made their decision. In contrast, the accuracy results tell us that the familiarity of features influenced the decision making of only the list-after participants. Thus, all participants may have had their search for information speeded or slowed by the presence of familiar features, but only list-after participants seemed to have had their decision making influenced by the familiarity of features. The use of a counting rule therefore seems to entail reliance on informational features, regardless of whether it is learned before or after perceptual experience of the categories. That is, counting rules seem necessarily to operate over abstract inputs. The ERP evidence, however, challenges this conclusion.

ERP activity

Given the exploratory nature of this research, we needed an objective means of selecting electrodes for analysis. Fortunately, just such an objective selection method is available via PLS (Lobaugh, West, & McIntosh, 2001; McIntosh et al., 1996), which does not require any a priori bias with respect to time-course or location of effects. PLS is similar to principle components analysis (PCA) in that it uses singular value decomposition to extract information from the dataset, but different in that it constrains analysis to the variance that can be explained by the experimental conditions. Singular value decomposition yields a set of latent variables (LVs; similar to eigenvalues in PCA) that represent particular contrasts, which account for a percentage of the cross-block covariance explained by the experimental conditions. Each singular value explains how much of the covariance was explained by a particular latent variable. One thousand permutations were computed and provided an estimate of obtaining a singular value by chance (similar to a p-value). The electrode saliences represent the relation between the experimental design contrasts (as represented by the LV) and the spatiotemporal pattern of ERP amplitude changes. Two hundred bootstrap re-samplings were performed to assess the reliability of electrode saliences at each time point by providing a standard error for each salience. The bootstrap procedure uses random sampling with replacement so that even though each sample will have the same number of elements as the original data, slightly different samples will be produced, and reliability of the saliences can be measured. The ratio of the salience to the standard error is approximately equal to a z-score, and so data points where the ratio was more than 1.7 ($p < .05$) were considered reliable.

Responses for the list-after participants were substantially faster than those for both count-after and count-before participants, therefore ERP analyses examined early and late windows. Two PLS analyses were performed, one including all rule groups over a 0–1000-ms time window and one including only the two counting groups over a 0–3000-ms window. Both analyses identified the same four regions of interest: frontal central sites (electrode FCz), central parietal sites (Pz), and two bilateral parietal sites, one on the left (CP3) and one on the right (CP4); these sites are shown in Figure 6. The early PLS analysis yielded one reliable LV discriminating groups and conditions, which accounted for 33% of the variance ($p < .05$). The late PLS analysis also yielded one reliable LV, which accounted for 66% of variance across groups, ($p < .003$).

Smaller epochs were created within each of the early and late PLS analyses. Epochs were chosen based on correspondence with the PLS and visual inspection. For the early time window, the first four epochs were 100 ms each (0–100, 100–200, 200–300, 300–400 ms), with two final epochs of 300 ms (400–700, 700–1000 ms). The late time window was divided into four epochs of 500 ms each (1000–1500, 1500–2000, 2000–2500, 2500–3000 ms). For each of these epochs we analysed amplitudes using 3 (group: list-after, count-before, count-after) × 3 (transfer item: facilitating familiarity, all novel, interfering familiarity) × 4 (location: FCz, CP3, CP4, Pz) repeated measures ANOVAs. As no meaningful effects of species (bleeb/ramus) emerged from the behavioural data, this factor was dropped for the amplitude analyses. Reliable effects were interpreted with post hoc Bonferroni-corrected t-tests.

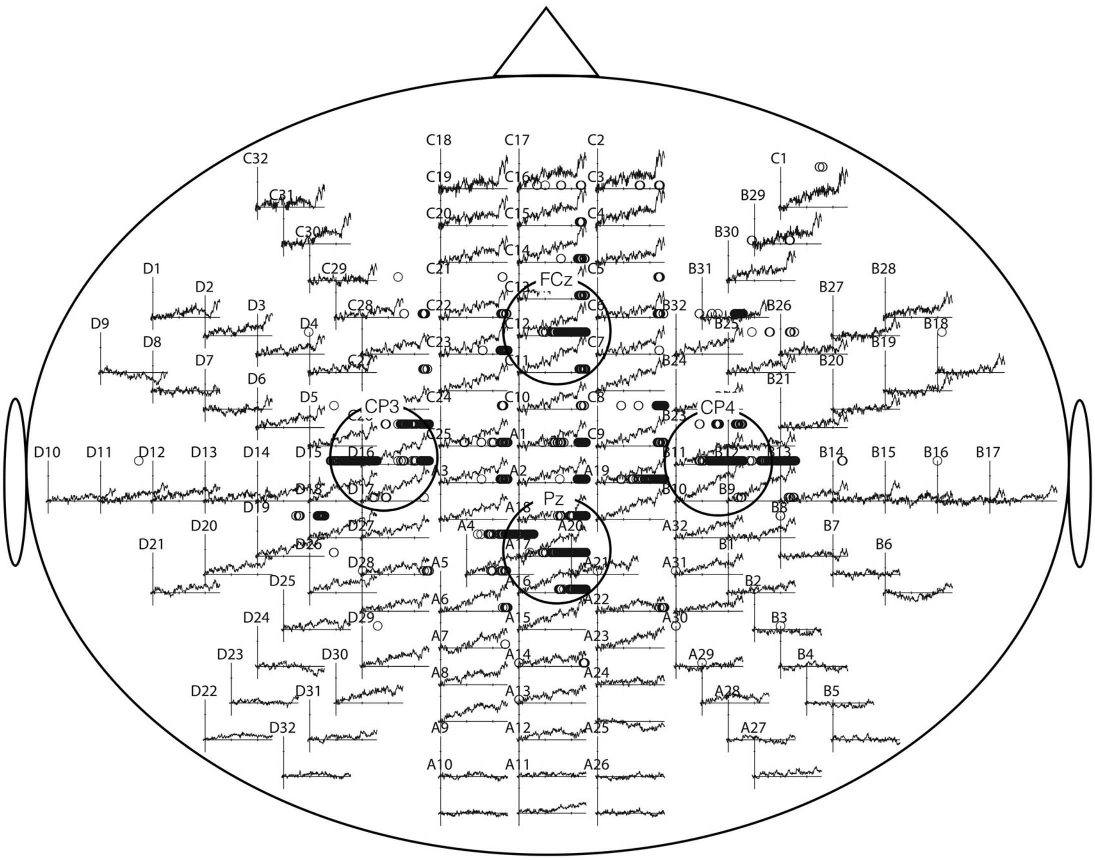

Figure 6. *Scalp locations of the four regions of interest (ROIs) identified by the late (0–3000-ms) partial least-square (PLS) analyses. The early PLS analysis produced similar regions of interest. The ROIs are centred on electrodes FCz, Pz, CP3, and CP4 (circled).*

Early analyses (0–1000 ms). Four of the six epochs yielded reliable effects, beginning as early as the 0–100-ms epoch; ANOVA results are summarized in Table 2. This is also the only epoch in the early window to reveal a main effect of transfer item. Bonferroni-corrected t-tests ($\alpha_{\text{adjusted}} = .017$) reveal that all-novel items produced a larger amplitude than facilitating-familiarity items, $t(44) = 3.49$, $SE = 0.154$, $p = .0024$, Hedges's $g = 0.50$. All rule groups seem to be responding to feature

Table 2. *Summary of ANOVA results for ERP amplitudes, early window*

Epoch (ms)	*Result*
0–100	Transfer item: $F(2, 90) = 4.11$, $MSE = 3.84$, $p = .02$, $\eta^2 = .084$
	Group × Location: $F(6, 135) = 2.68$, $MSE = 3.14$, $p = .02$, $\eta^2 = .106$
100–200	Group × Location: $F(6, 135) = 2.32$, $MSE = 5.78$, $p = .04$, $\eta^2 = .093$
400–700	Group × Location: $F(6, 135) = 2.53$, $MSE = 12.51$, $p = .02$, $\eta^2 = .101$
700–1000	Group × Location: $F(6, 135) = 3.63$, $MSE = 14.81$, $p = .002$, $\eta^2 = .139$

Note: ANOVA = analysis of variance; ERP = event-related potential.

familiarity, as there was no Group × Transfer Item interaction.

The only other effect found across early epochs was an interaction of Group × Location. The waveforms in Figure 7 show the count-before group with a larger amplitude than either the count-after or list-after groups at site FCz; however, neither of the Bonferroni-corrected *t*-tests proves reliable. For the 100–200-ms epoch, the count-before group appears to have a larger amplitude than the list-after group at CP3, but the difference falls short of reliability after Bonferroni correction.

Inspection of the waveforms in Figure 7 also reveals striking differences between groups at the Pz site, beginning at 400 ms. Although none of the Bonferroni-corrected *t*-tests yields a reliable result ($\alpha_{adjusted} = .00075$),[3] the Group × Location interaction tells us that the groups differ at at least one location, and the largest group differences in this time period are between the count-before group and the other two at the Pz site (count-

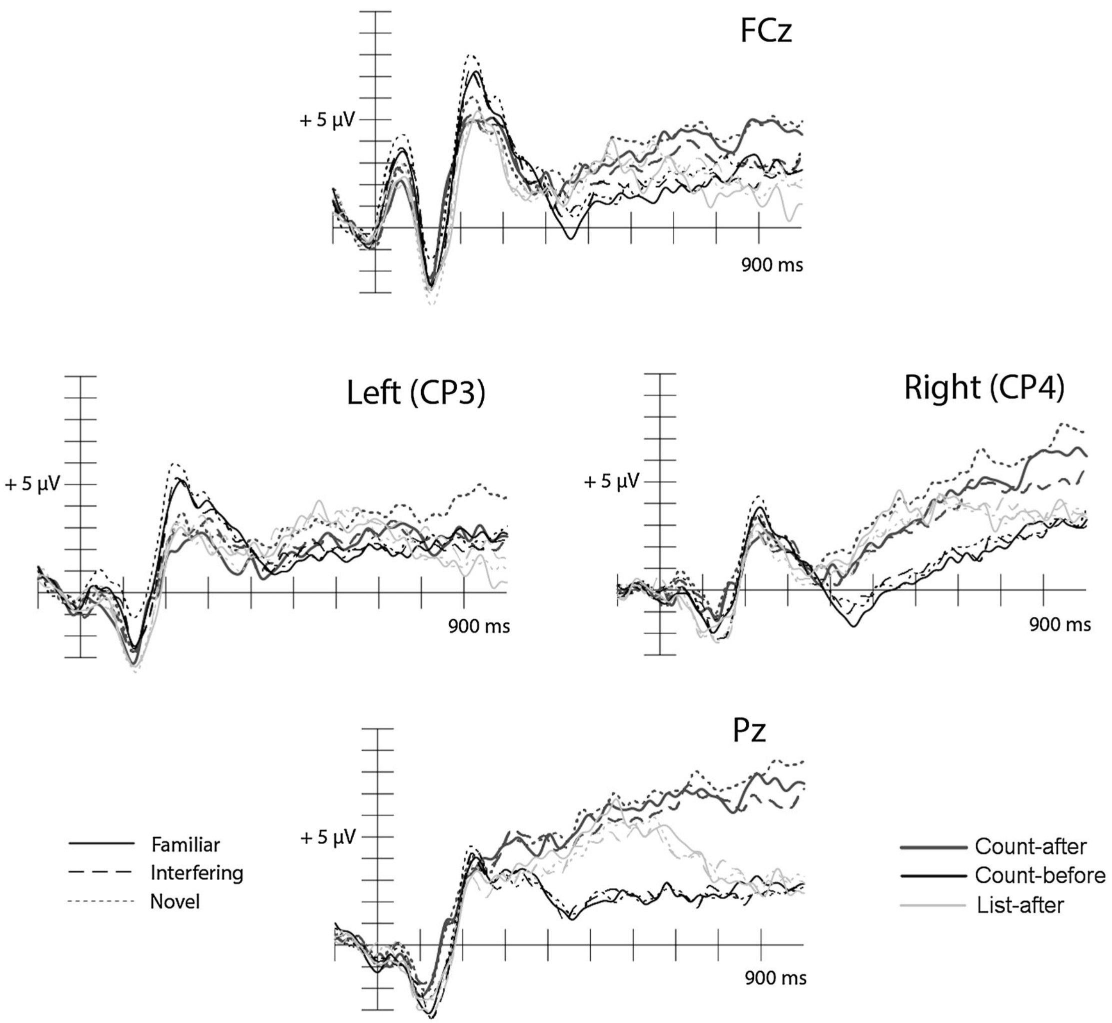

Figure 7. *Waveforms appearing at electrodes FCz, Pz, CP3, and CP4 during the 0–1000-ms time window. Amplitude differences were analysed at six epochs: 0–100 ms, 100–200 ms, 200–300 ms, 300–400 ms, 400–700 ms, and 700–1000 ms.*

before vs. count-after, Hedges's $g = 0.59$; count-before, vs. list-after, Hedges's $g = 0.47$; count-after vs. list-after, Hedges's $g = 0.13$). The activity at the CP4 site is similar in pattern, although the effects are smaller (count-before vs. count-after, Hedges's $g = 0.32$; count-before, vs. list-after, Hedges's $g = 0.45$; count-after vs. list-after, Hedges's $g = 0.03$). In contrast, the largest effect at the other two sites is the count-before versus list-after contrast at the CP3 site, and this is moderate at best, Hedges's $g = 0.23$. Thus, the reliable interaction seems to be driven by activity at the Pz and maybe CP4 regions, where the contrast seems to be between the count-before group and the two "after" groups, rather than between the counting-rule and feature-list group.

By 700 ms, however, the list-after group amplitudes drop down to those of the count-before group, as if some processing had been completed. In this final window, the count-after group produced greater amplitudes than either the count-before group, $t(30) = 3.95$, $SE = 1.26$, $p = .0004$, Hedges's $g = 0.70$, or the list-after group, $t(30) = 3.95$, $SE = 1.26$, $p = .0004$, Hedges's $g = 0.70$.

The early activity suggests that all groups initially register the differences in feature familiarity very early on, perhaps reflecting an advantage for processing or attending to familiar perceptual forms as indexed by the lower amplitudes for both facilitating- and interfering-familiarity transfer items. However, as processing progresses, the groups begin to pull apart. The list-after and count-after groups initially look more different from the count-before group, possibly because they are processing instantiated features and thus are interpreting/identifying richer feature representations, which takes more processing effort and yields a more positive amplitude. This process appears to end earlier for the list-after group, possibly because the search ends after identifying only one or two features; from about 700 ms the list-after group amplitudes drop down to those of the count-before group. In contrast, the amplitudes of the count-after group continue to climb as they seek out two or three features, according to the new rule they have been given. At a minimum, these results suggest that the two counting-rule groups show qualitatively different forms of ERP activity at the midline parietal site, despite a common rule and nearly identical behavioural responses. The differences between counting-rule groups grow more robust during the late window.

Late analyses (1000–3000 ms). During the first epoch (1000–1500 ms), only the Group × Location interaction is reliable. From 1500 ms on, all epochs yield the same two reliable effects, consisting of a Group × Test Item interaction and a Group × Location interaction. ERP responses are rarely examined at such late windows, and some caution in interpreting them is necessary; however, the results across the late epochs are quite consistent. ANOVA results for all late epochs are summarized in Table 3.

In all epochs, the Group × Location interaction involved a trend in which the count-after group had a more positive amplitude than the count-before group at CP4 and Pz sites. However, the Bonferroni-corrected $\alpha = .00075$, and, thus, none of our paired t-tests achieve significance, although three comparisons at the Pz electrode approach significance ($p < .0015$): 1500–2000 ms, $t(30) = 3.51$, $SE = 2.245$, $p = .0014$, Hedges's $g = 0.62$; 2000–2500 ms, $t(30) = 3.52$, $SE = 2.783$, $p = .0014$, Hedges's $g = 0.62$; 2500–3000 ms, $t(30) = 3.59$, $SE = 3.305$, $p = .0012$, Hedges's $g = 0.63$.

For all Group × Test Item interactions, the interactions resulted from the all-novel transfer items producing higher amplitudes than the interfering-familiarity transfer items, but only for the count-after group ($\alpha_{adjusted} = .00139$[4]). For the 1500–2000-ms epoch: all-novel count-after > interfering-familiarity count-after, $t(15) = 3.94$, $SE = 1.04$, $p = .00130$, Hedges's $g = 0.70$. For the 2500–3000-ms epoch: all-novel count-after > interfering-familiarity count-after, $t(15) = 4.38$,

[3]However, with the interaction involving 12 (4 × 3) means, and thus (12 × 11)/2 = 66 possible post hoc comparisons, the Bonferroni correction becomes very conservative.

[4]$\alpha_{adjusted} = .05/36$, given 36 possible comparisons based on nine (3 × 3) means.

Table 3. *Summary of ANOVA results for ERP amplitudes, later window*

Epoch (ms)	*Result*
1000–1500	Group × Location: $F(3, 90) = 5.71$, $MSE = 31.79$, $p = .001$, $\eta^2 = .160$
1500–2000	Group × Item: $F(2, 60) = 4.84$, $MSE = 26.67$, $p = .011$, $\eta^2 = .139$
	Group × Location: $F(3, 90) = 7.02$, $MSE = 53.76$, $p < .001$, $\eta^2 = .190$
2000–2500	Group × Item: $F(2, 60) = 3.97$, $MSE = 32.71$, $p = .025$, $\eta^2 = .117$
	Group × Location: $F(3, 90) = 7.76$, $MSE = 80.93$, $p < .001$, $\eta^2 = .205$
2500–3000	Group × Item: $F(2, 60) = 5.43$, $MSE = 36.52$, $p = .007$, $\eta^2 = .153$
	Group × Location: $F(3, 90) = 8.22$, $MSE = 114.51$, $p < .001$, $\eta^2 = .215$

Note: ANOVA = analysis of variance; ERP = event-related potential.

$SE = 1.25$, $p = .0005$, Hedges's $g = 0.77$. Although the trend is for the all-novel count-after > interfering-familiarity count-after during the 2000–2500-ms epoch, this only approaches significance ($p < .0028$), $t(15) = 3.75$, $SE = 1.228$, $p = .0019$, Hedges's $g = 0.66$.

As the waveforms shown in Figure 8 suggest, striking differences between the two counting groups emerge from 1000 ms onwards, with the count-after group maintaining the enhanced positivity relative to the count-before group demonstrated earlier at the Pz site. Perhaps most relevant for the issue of feature representation, however, is the finding that only the count-after group showed a sensitivity to the familiarity of the transfer items' feature forms.

Summary of results

Overall, the activity of the list-after and count-after groups suggests that they were reliant on features that require more processing than those relied on by the count-before group, consistent with the proposition that the count-after and list-after groups relied on instantiated features to make classification decisions while the count-before group relied on informational features. This interpretation would also imply that informational and instantiated features are qualitatively distinct. Further, the late sensitivity to feature familiarity that emerges for the count-after group suggests that the perceptual details influence processing even during the late decision-making stage. This is consistent with the proposition that the count-after participants were applying their rule to instantiated, rather than informational, features.

Discussion

We set out with this study to try to refine our understanding of instantiated and informational features, and the relations between feature abstraction and rules. We wanted to know whether, as implied by earlier treatments (e.g., Brooks & Hannah, 2006; Hannah & Brooks, 2006, 2009), informational and instantiated features were qualitatively different from one another, with each type of feature representation conveying unique information. We wanted to know whether strong rules such as counting rules necessarily require (abstract) informational features as inputs, or whether they can operate across instantiated features. How much of categorization behaviour reflects the nature of the rules used, and how much reflects the constraints of the representations the rules act on?

The behavioural evidence suggested that the count-before and count-after groups are engaged in similar processing, which is distinct from that of the list-after group. The two counting-rule groups do not look the same in the ERP data, however. Instead, from 400 ms onwards, the count-after group looks more like the list-after group than the count-before group; from 700 ms onwards, the count-after group is still different from the count-before group, while the list-after group resembles the count-before group. Over the

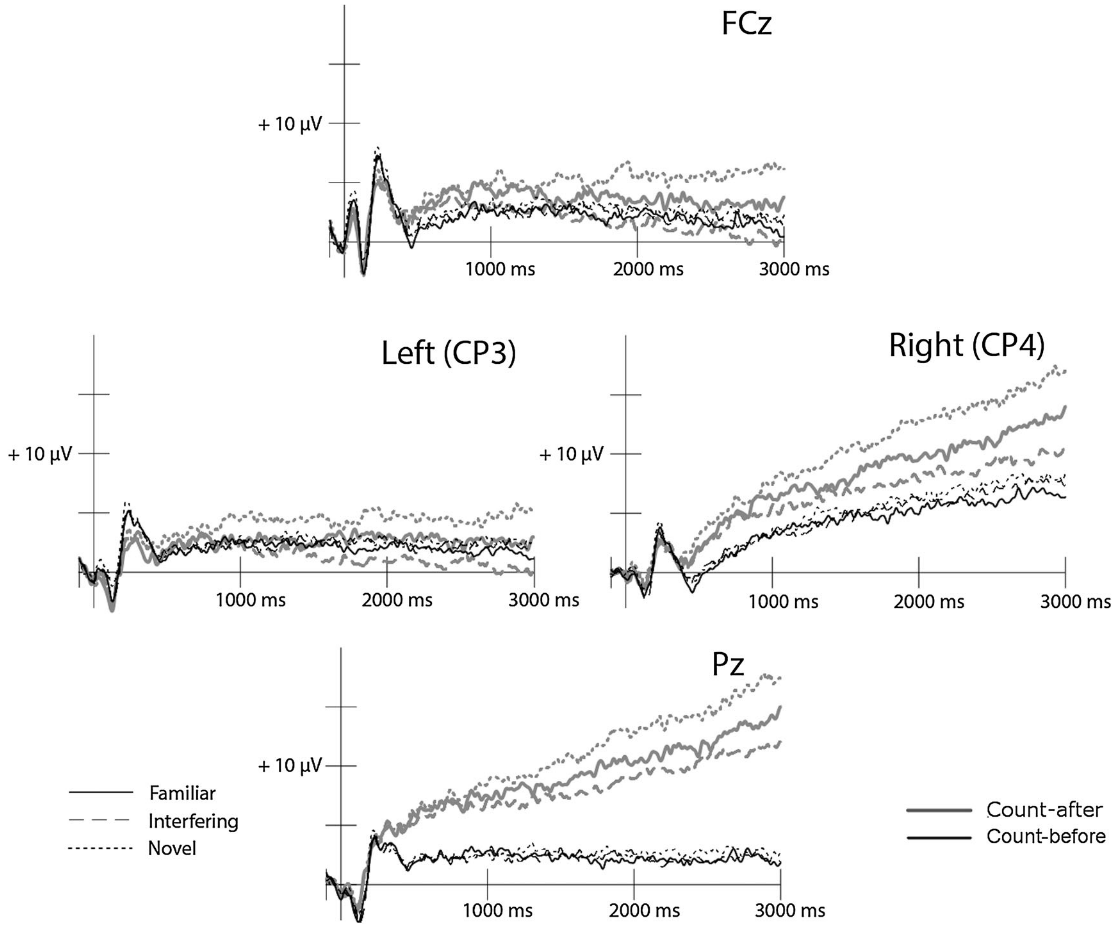

Figure 8. *Waveforms appearing at electrodes FCz, Pz, CP3, and CP4 during the 0–3000-ms time window. Amplitude differences were analysed at four epochs: 1000–1500 ms, 1500–2000 ms, 2000–2500 ms, and 2500–3000 ms. Only the two counting-rule groups were analysed in this window.*

later epochs, only the ERPs of the count-after group show differences between familiar and novel feature forms.

Despite their similar behavioural profile and the use of an identical rule, the two counting-rule groups yielded distinct neural signatures, particularly at intermediate and later time points most strongly correlated with likely decision processes. The ERP patterns suggest the following. First, the two counting-rule groups are relying on different sources of feature information. The count-after group—which shows prolonged elevated positivity—appears to rely on processing that is more demanding. This would be consistent with the count-after group relying on feature representations that were more detailed, and carrying more information, than the count-before group. That is, the count-after participants are applying their rule over instantiated features while the count-before participants apply the same rule to informational features. This conclusion is strengthened when we consider the initial similarity between the list-after and count-after profiles. Second, the distinctness of the ERP profiles of the count-before and two rule-after groups implies that instantiated and informational features are qualitatively different from one another: Informational features are not sparse instantiated features, but convey information not captured by instantiated counterparts.

All groups show very early sensitivity to the familiarity of the specific feature forms of the transfer items. This is consistent with instantiated features influencing attention or visual search before any decision process begins, as Hannah and Brooks (2009) proposed. Thus, even strong rule users applying their rule to informational features may have their implementation of this rule influenced by instantiated features. Familiar-looking features may get evaluated first, and this may, under normal conditions, make rule use more efficient.

Feature representation

Although the exploratory nature of this work precluded precise hypothesizing about the nature of the waveform differences that we could expect, the waveform differences seen at the Pz/CP4 sites (midline parietal and right-hemisphere sites) are consistent with the idea that count-before participants relied on informational features while count-after and list-after participants incorporated instantiated features into their decision making. We offer the following observations and interpretations as a preliminary account of the ERP activity.

At the Pz/CP4 clusters, the count- and list-after groups seemed to separate from the count-before group, as supported by the Group × Location interactions. The separation was especially apparent at the Pz sites, particularly during 400–700 ms after stimulus presentation. During this window, ERP activity began to return to baseline for the count-before group, while amplitudes steadily increased for both count-after and list-after groups. This pattern of rising activity for the rule-after groups and falling activity for the count-before group is consistent with the idea that the midline activity reflects the effort of processing the feature representations—the more complex the feature, the greater the response. If this interpretation is correct, then the rule-after groups seem to be processing feature representations that require more effort than those processed by members of the count-before group. The Pz activity is thus consistent with the count-before participant processing abstract informational features, while members of the other two rule groups are processing detailed instantiated features.

An alternative interpretation is possible. At the start of transfer, both rule-after groups have had 80 fewer trials to practise their rule than did the count-before group. The count-before group, therefore, may have simply been more practised at retrieving and using their rule than either rule-after group. Repetition of stimuli has long been known to attenuate neural responses (see Grill-Spector, Henson, & Martin, 2006, for a concise review). Practice of a task, such as discriminating two sounds (e.g., Reinke, He, Wang, & Alain, 2003), also attenuates ERPs. As the list-after group presumably makes their decision after processing only one or two features, they maintain their rule for a shorter time, producing an earlier return to baseline than the count-after group.

This explanation cannot be ruled out, but is less plausible than the representational argument given above. Neural responses to practice or repetition are usually correlated with performance changes. Thus, if the count-before group were simply more efficient at rule use than count-after participants, then we would expect them to be faster than the count-after participants in making a decision. However, this was not the case.

Fortunately, both the postulate that instantiated and informational features are qualitatively different and the postulate that strong rules can accept instantiated inputs are supported by other data, particularly the striking late differences between counting groups over the right central parietal area. Over the 1500–3000-ms window, the count-after group's ERPs were sensitive to feature familiarity, but the count-before group did not show this sensitivity. There are intriguing commonalities in the late waveforms as well, with both groups showing a slow rising increase in positivity over the right central parietal area. Such commonalities could be indexing the same decision process—that is, the use of the same counting rule. The patterns seem readily understandable if they reflect that the count-before participants were applying their rule to informational features, while the count-after participants applied the same rule to instantiated features.

Any explanation of the differences between the count-after and count-before groups is necessarily speculative. What is not speculative, however, is that both counting-rule groups (a) used the same rule, (b) trained on the same stimuli, and (c) showed the same behavioural pattern. Given these constraints, we suggest that the differences between the groups' waveforms are most plausibly consistent with the two groups relying on different types of feature representations.

Informational and instantiated features

Informational features not simply seem to be less detailed, or fuzzier, representations of some specific feature form, but contain information that is not found in any specific feature. How then do informational features form out of instantiated features? To what extent is such feature construction a deliberate process? That informational features contain generic information distinct from that conveyed by instantiated features suggests that they emerge from some sort of comparison process. This in turn implies some sort of deliberative processing, even if such abstraction is incidental to some focal task (e.g., Whittlesea & Dorken, 1993). If so, such a deliberative origin would seem to speak against their emergence as a result of automatic abstraction processes, such as those claimed in fuzzy-trace theory (Brainerd & Reyna, 1990a, 1990b, 1996; Reyna & Brainerd, 1995), or implicit memory accounts (e.g., Knowlton, Ramus, & Squire, 1992).

Another possibility is that informational features are Barsalou's (1999, 2009) simulators, at the level of features. This would lead to a picture of feature extraction by the registration of commonalities across features by conjunction neurons in association areas. However, while Barsalou has argued that such extraction requires attention, it is not clear whether this requires explicit registrations of such commonalities.

Gentner and Medina (1998) described rule abstraction as a process of deliberate comparison across aligned exemplars. Such an aligned comparison could be extended to features, such that we could think of informational features as feature rules. Yamauchi (2009) has recently stressed the importance of labels in guiding such alignment-based abstraction. The feature labels contained in the counting rule provided to the count-before participants would allow them to generate a generic representation of "two legs", "striped coat", and so on, from memory, and these initial informational features then guide the count-before participants' analysis and encoding of the training features. Brooks and Hannah (2006) similarly found that participants given a feature list prior to training showed much reduced sensitivity to the familiarity of features, suggesting a greater reliance on informational features.

Features and tags

One reviewer, however, suggested an alternative that merits consideration: All feature representations are instantiated. In this view, people recruiting a collection of feature instances and interpreting new features based upon similarity to features in the collection can explain the apparent generalization of feature knowledge. Such a view implies that features are themselves categories (Schyns, Goldstone, & Thibaut, 1998), and that feature categories consist exclusively of feature instances. This instance-theory approach (Brooks, 1978) has appeal to us, but it cannot explain the ERP data presented here without introducing a de facto informational representation.

The reviewer argued that observed differences between counting groups in our experiment may arise because count-before participants hierarchically organize instantiated features acquired in training around a verbal label, or node or a representation that integrates features that share some property, building up feature categories. Count-after participants, in contrast, have a "flatter" organization, with feature instances directly connected to a category label; the count-after participants may gradually build up such a hierarchical organization as a result of applying the counting rule over the transfer period. The counting rule for count-before is applied to the node/label itself. The problem for such an account is that the node/label representation is, on its own, an informational feature: It represents a collection of feature instances by association with them, but does not itself carry any specific form information.

However, although this account does not escape the informational/instantiated distinction, it has at its heart an idea that allows an elegant and precise explanation of their relation. To make this leap, we also need to invoke Clark's (2008) notion of words as tags to experiences or ideas. A feature name can be used to tag experience, such as the experience of a particular feature form. Critically, we suggest that a tag can be used in one of two ways: It can act as a *retrieval cue* activating its associated experience, or it can act as a *substitute* for the tagged experience. Used to retrieve feature instances, it produces instantiated effects; used as a substitute for feature instances, it produces informational effects. Such an account gives us a precise definition of informational and instantiated features and also makes it easy to understand how people can be flexible in the information they recruit, simply by shifting how the tag is used. It further suggests how informational features emerge—not from a representational change, but from a new way of using the existing representations, by confining processing to just the tag.

Rules and representations

Our results also suggest that strong rules can operate over either instantiated or informational features. That a rule can take highly specific, detailed representations as inputs seems in conflict with traditional approaches to rules as abstract processes that operate over abstract features (e.g., Sloman, 1996). The finding also sits uncomfortably with a dichotomy splitting cognitive processes into rule-based versus similarity-based, where the former is a marker of abstract processes and the latter a marker of embodied or instantiated processes.

This rules-versus-similarity dichotomy has guided much cognitive research, especially in the domain of categorization, even as it has come under increasing criticism (e.g., Pothos, 2005). For example, Folstein and Van Petten (2004) conducted an ERP-based study of concept learning in which half of the participants were taught a decision rule, and half were not. The supposition was that this second group would rely on similarity, and that this would be a qualitatively different process from that used by the rules group. However, both groups showed a similar response to repeated test features, suggesting a common response to feature form for both strategy groups, despite Folstein and Van Patten's claims that the ERP work supported such a rules/similarity distinction.

However, our results also qualify Hannah and Brooks' (2009) claim that strong rules are linked to informational features, and weak rules to instantiated features, neglecting the issue of the necessity of the linking of the strong rules and informational features. This work fills in that gap and points to a merely contingent relation between strong rules and informational features: Informational features are more likely to overlap across categories, yielding items with conflicting information. Thus, encoding items in terms of their informational features is more likely to produce a situation where a rule that resolves such conflict is necessary. The nature of the relation between weak rules and instantiated features, however, is not addressed by our present results.

REFERENCES

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289. Retrieved from http://doi.org/10.1098/rstb.2008.0319

Brainerd, C. J., & Reyna, V. F. (1990a). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*, 3–47.

Brainerd, C. J., & Reyna, V. F. (1990b). Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review*, *10*, 365–403.

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology*, *32*, 467–478.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd, Barbara (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Lawrence Erlbaum.

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules". *Journal of Experimental Psychology: General*, *135*, 133–151. Retrieved from http://doi.org/10.1037/0096–3445.135.2.133

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Folstein, J. R., & Van Petten, C. (2004). Multidimensional rule, unidimensional rule, and similarity strategies in categorization: Event-related brain potential correlates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1026–1044. Retrieved from http://doi.org/10.1037/0278-7393.30.5.1026

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*, 45–56.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*, 263–297.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10* (1), 14–23. Retrieved from http://doi.org/10.1016/j.tics.2005.11.006

Hannah, S. D., & Brooks, L. R. (2006). Producing biased diagnoses with unambiguous stimuli: The importance of feature instantiations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1416–1423. Retrieved from http://doi.org/10.1037/0278-7393.32.6.1416

Hannah, S. D., & Brooks, L. R. (2009). Featuring familiarity: How a familiar feature instantiation influences categorization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *63*, 263–275. Retrieved from http://doi.org/10.1037/a0017919

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*, 172–179.

Lobaugh, N. J., West, R., & McIntosh, A. R. (2001). Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology*, *38*, 517–530. Retrieved from http://doi.org/10.1017/S0048577201991681

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. *Psychological Science*, *8*, 363–367.

Markman, A. B., & Maddox, W. T. (2003). Classification of exemplars with single- and multiple-feature manifestations: The effects of relevant dimension variation and category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 107–117.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, *3*, 143–157. Retrieved from http://doi.org/10.1006/nimg.1996.0016

Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 607–625.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences*, *28*, 1–14. Retrieved from http://doi.org/10.1017/S0140525x05000014

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General*, *122*, 92–114.

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research*, *17*(3), 781–791. Retrieved from http://doi.org/10.1016/S0926-6410(03)00202-7

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Schyns, P. G., Goldstone, R. L., & Thibaut, J-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sweller, N., & Hayes, B. K. (2010). More than one kind of inference: Re-examining what's learned in feature inference and classification. *The Quarterly Journal of Experimental Psychology*, *63*, 1568–1589. Retrieved from http://doi.org/10.1080/17470210903438547

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A*, *47*, 631–650. Retrieved from http://doi.org/10.1080/14640749408401131

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*, 227–248.

Yamauchi, T. (2009). Finding abstract commonalties of category members. *Journal of Experimental & Theoretical Artificial Intelligence*, *21*, 155–180. Retrieved from http://doi.org/10.1080/09528130802113299

APPENDIX A

Debriefing protocol

Debriefing of participants began with an open-ended question: "How did you decide whether something was a bleeb or ramus?" If this produced a counting rule as an answer, or another strong rule, questioning stopped. Otherwise, this was followed by a more specific question involving one of the named features to determine whether they were aware that no single feature was sufficient for categorization. For example, if a counting participant said, "I used mainly legs and bodies to make my decisions", the experimenter would probe with: "If an animal had two legs [or, rounded torso . . .], what would you call it?" If the participant replied with "bleeb", this would be indicative of a feature-list rule. If their response indicated that a single feature was not enough (e.g., "Well. That would depend on whether the torso was rounded or not"), then this would be indicative of a counting rule. If their strategy was still not clear the experimenter probed with a more specific question: "Could you ever make a decision just by looking at the legs alone or body alone?" If the participant answered in the positive (e.g., "sometimes") then they were counted as using a list rule. If they answered in the negative, then they were counted as using a counting rule.

APPENDIX B

Behavioural analysis of variance (ANOVA) results involving species

Accuracy

The species factor yielded only two effects in the accuracy data, both two-way interactions. There was a Transfer Item × Species interaction, $F(2, 90) = 5.59$, $MSE = .002$, $p = .005$, , $\eta_p^2 = .11$, that reflected a larger effect of transfer item on bleebs than ramuses. The Rule Group × Species interaction, $F(2, 45) = 4.32$, $MSE = .040$, $p = .019$, $\eta_p^2 = .16$, is explained by the observation that both counting groups were more accurate for ramuses than for bleebs, while the reverse held true of the listing participants.

Reaction time

Bleebs were correctly identified more quickly than ramuses, $F(1, 45) = 17.49$, $MSE = 313{,}191$, $p < .001$, $\eta_p^2 = .28$. This species main effect appeared larger for the two counting groups than for the listing group, and the Rule Group × Species interaction was reliable, $F(2, 45) = 5.90$, $MSE = 313{,}191$, $p = .005$, $\eta_p^2 = .19$. The species difference also appeared larger for the all-novel condition than for the interfering- and facilitating-familiarity conditions, and the Transfer Item × Species interaction proved reliable, $F(2, 90) = 3.66$, $MSE = 60{,}554$, $p = .030$, $\eta_p^2 = .08$.