The arbitrariness of the sign:

Learning advantages from the structure of the vocabulary

Padraic Monaghan[1], Morten H. Christiansen[2], & Stanka A. Fitneva[3]

[1]Department of Psychology & Centre for Research in Human Development and Learning,

Lancaster University, UK

[2]Department of Psychology, Cornell University, USA

[3]Department of Psychology, Queen's University, Canada

in press, *Journal of Experimental Psychology: General*

Corresponding author:

Padraic Monaghan

Department of Psychology

Lancaster University

Lancaster

LA1 4YF

UK

E-mail: p.monaghan@lancaster.ac.uk

Tel: +44 1524 593813

Fax: +44 1524 593744

Abstract

Recent research has demonstrated that systematic mappings between phonological word forms and their meanings can facilitate language learning (e.g., in the form of sound symbolism or cues to grammatical categories). Yet, paradoxically from a learning viewpoint, most words have an arbitrary form-meaning mapping. We hypothesized that this paradox may reflect a division of labor between two different language learning functions: arbitrariness facilitates learning specific word meanings and systematicity facilitates learning to group words into categories. In a series of computational investigations and artificial language learning studies, we varied the extent to which the language was arbitrary or systematic. For both the simulations and the behavioral studies, we found that the optimal structure of the vocabulary for learning incorporated this division of labor. Corpus analyses of English and French indicate that these predicted patterns are also found in natural languages.

Introduction

From the sound of a word alone, there is typically little information available about its meaning. This arbitrariness of the sign – where there is minimal systematicity between the forms and meanings for words in the vocabulary – is a well-established property of the world's languages (de Saussure, 1916; Hockett, 1960). However, from a language learning perspective this property presents a challenge, because the set of mappings between words and referents that have already been acquired by the learner will not assist in deducing the meaning of a novel item from its phonology and prosody alone. The arbitrariness of form-meaning mappings thus presents a problem for learning because the resources required to form an arbitrary mapping are greater than those required for a systematic mapping. So, why is arbitrariness not only tolerated in the vocabulary, but inherent within language structure to the point that it is considered a linguistic "universal feature" (Greenberg, 1957)?

The prevalence of arbitrary form-meaning mappings thus seems counterintuitive to theories of learning. Indeed, Western scholars have for centuries found offense in the arbitrariness of the form-meaning mapping, and consequently sought to discover the perfect language — the biblical "universal language" existing prior to the fall of the tower of Babel — which was assumed to incorporate a high degree of systematicity instead of the abundant arbitrariness characteristic of modern languages (see Eco, 1995, for a review). To this end, Wilkins (1668) wrote "An Essay towards a Real Character and a Philosophical Language" in which he proposed a vocabulary that systematically related certain letters to certain meaning classes. The first two letters distinguished 40 genera of concepts, and the remainder of the word, letter by letter, provided nested sub-category information. But why do natural languages not follow Wilkins' proposal for a perfect language? That is, if arbitrary mappings between form and meaning present a substantial cost in terms of learning, compared to systematic ones, because they require each word to be learned independently, then why do we *not* find that languages with systematic mappings have been selected for in the process of language evolution (Christiansen & Chater, 2008)?

Wilkins may inadvertently have provided part of the answer to these questions in his own work. As noted in Borges' (1942) critical essay on Wilkins' language, there were difficulties in determining a classification system that would easily accommodate every known concept. Hence, classifications were forced for many concepts and did not correspond to "natural" categories. A whale, for instance, was classified as an "oblong viviparous fish", and a dog was a "rapacious viviparous animal with an oblong head", descriptions which are unlikely to occur as spontaneously produced properties of these concepts (e.g., Nelson, McEvoy, & Schreiber, 2004). In addition, Wilkins' "perfect", but inflexible, classification system is further undermined by the tendency of concepts to change considerably both synchronically (according to context) as well as diachronically (across the life-span and generational time).

The problems with Wilkins' attempt at a perfect language, however, can only provide negative clues to why languages are not as systematic in their form-meaning mappings as might be expected based on learning theory. To explain why arbitrary mappings are so prevalent it would seem that evidence of some sort of benefit from such mappings is required. Here we consider whether arbitrariness in form-meaning mappings might provide, in some way, an advantage for language learning that serves to compensate for the disadvantage of both the greater resources required for forming the mapping and the absence of generalisation from previously acquired vocabulary. This is not a new theoretical proposal (see, e.g., Eco, 1995), yet, to our knowledge, no empirical study of this hypothesis has been conducted.

In this paper, we present a series of computational and experimental studies to address the effect of form-meaning arbitrariness in learning. The computational models explore the extent to which the arbitrary advantage is a consequence of general learning principles that become embodied within the vocabulary structure. The hypotheses generated by the computational modeling are then tested using artificial language learning experiments that either incorporate arbitrariness or systematicity into the language. However, these empirical studies of arbitrariness in vocabulary learning reveal a tension between two communicative tasks that are required in language learning and comprehension. We show that the task of *individuating* particular referents of words is promoted by arbitrariness, but the task of *categorizing* sets of words according to their semantic features into their respective grammatical classes is impeded by arbitrariness. We also show that the arbitrariness advantage occurs only under certain circumstances of learning, which raises testable hypotheses for the extent to which arbitrariness is observed in natural languages. The experimental and modeling results combine to suggest that a *division of labor* is expressed within the structure of the form-meaning mappings in natural languages: arbitrariness supports meaning individuation for individual words and systematicity supports the grouping of words into categories. Finally, we report data from corpus analyses of the vocabulary of English and French to indicate how the structure of words in these two languages incorporates such a division of labor between arbitrariness and systematicity.

In the remainder of the Introduction, we review the extent to which the vocabulary of natural languages actually is arbitrary. We then review previous explanations for arbitrariness in terms of how it may provide an advantage for language learning, before presenting our modeling and experimental studies.

How arbitrary is the vocabulary?

The conventional view of the structure of the vocabulary is that form-meaning mappings are entirely arbitrary (de Saussure, 1916; Greenberg, 1957). However, there are several exceptions to this where small pockets of systematicity are detectable in the vocabulary. These have attracted considerable interest and research effort, principally because they stand as exceptions to the general arbitrariness of the vocabulary.

Sound-symbolism is one such indication of systematicity. Onomatopaeia (Langdon, 1984) and phonaesthemes – particular phonemes or phoneme clusters that reflect similarities between word meanings (e.g., in English, words beginning *gl*- often refer to light, Bergen, 2004, p.293) – have been shown in some cases to be statistically reliable in corpus analyses (Otis & Sagi, 2008; Tamariz, 2008). Though phonaesthemes are language-specific, other sound-symbolic properties are proposed to be language-general. Expressives referring to large sizes containing low vowels (*large*/*humungous*), and those referring to small sizes containing high vowels (such as *tiny*/*little* (Hinton, Nichols, & Ohala, 1994; Ohala, 1984). As another example of cross-linguistic generalities, in a study of 136 languages, Ultan (1978) found that all languages that made a near versus far deictic distinction, also formed this distinction in terms of a high/front versus low/back vowel, respectively (e.g., *this/that* in English). Furthermore, synaesthetic relationships between certain sounds and meanings provide more evidence of systematicity in the vocabulary (Kovic, Plunkett, & Westermann, 2010; Nygaard, Cook, & Namy, 2009; Spector & Maurer, 2009).

These instances of systematicity in word-meaning mappings indicate that there are exceptions to the arbitrariness of the language, though this is only for small and specific subsets of the language. Each systematic relationship relates to some isolated semantic feature distinction, such as light-emittance (*gl-*), mass (high/low vowel), or proximity in deixis (back/front vowel). However, there is the possibility that certain semantic features are symbolised in phonology more systematically. In particular, semantic features that relate to grammatical category distinctions (Pinker, 1984) can be reflected in the morphological structure of the language, such as the action/object distinction. Thus, in order to acquire the language, the child is required to learn the categories to which these semantic features relate. Learning similarities among referents is facilitated by similarities among the form of the words referring to them. Cassidy and Kelly (1991, 2001) and Fitneva, Christiansen, and Monaghan (2009) found that children learn object and action referents better if there is a correspondence between the sounds of the nouns and verbs and their respective grammatical categories. For learning artificial categories, too, acquisition was facilitated if there was phonological coherence among the words referring to the different categories (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Monaghan, Chater, & Christiansen, 2005).

Morphology can therefore be effective and prevalent in reflecting semantic feature distinctions. Most obviously, morphology can instantiate iconic relationships between form and meaning: Plurals tend to be marked forms (Bybee, 1985), expressed in English by adding an inflectional morpheme, though in other languages it may be expressed through reduplication (as in Somali: dab/dabab (gloss: fire/fires), tug/tugag (gloss: thief/thieves), Greenberg, 1957). The greater length of the word reflects the greater number or size of the referent (Newmeyer, 1998). But morphology also provides many instances of systematicity that are not iconic, as in the relations between morphemes and grammatical categories, which may contribute significantly to effective learning of categories (St Clair, Monaghan, & Ramscar, 2009). In English, endings such as *–ing*, which usually indicate verbs suggest that the meaning of the word may entail an action or an activity, and the ending *–ly* generally ends an adverb, indicating that the word refers to manner of the action (Fudge, 1984).

Systematicity at the grammatical category level can also be observed in terms of allophonic, phonological and prosodic variation. Kelly (1992) summarised a range of phonological and prosodic cues that reflected grammatical category distinctions in English. More recently, Monaghan, Christiansen, and Chater (2007) showed that there were multiple phonological cues applying in a range of languages (across both Indo-European and Japonic families), and that, taken together, these cues could effectively distinguish grammatical categories of words in each vocabulary to a high degree of accuracy. In a recent count for English, 21 phonological or prosodic cues were found to reflect grammatical categories (Monaghan & Christiansen, 2008), indicating that systematicity is widespread and substantial for distinguishing grammatical categories (see also Farmer, Christiansen, & Monaghan, 2006).

In sum, the vocabulary appears to be almost entirely arbitrary, with small pockets of systematicity in terms of sound-symbolism for specific subgroups of words, such as expressives, deictic pronouns, or clusters of phonaesthemes. But there does appear to be substantial systematicity in terms of semantic features that relate to grammatical categories, as reflected in sound-category correspondences.

Potential explanations for the arbitrary advantage

Research on language evolution has indicated that many aspects of linguistic structure may be explained by fast-working cultural transmission mechanisms rather than slower changing biological adaptations (for a review, see Christiansen & Chater, 2008). These theories of the cultural evolution of language emphasize ease of learning and processing as important factors in the survival of various components of language. As noted by Darwin (1874, p. 91): "A struggle for life is constantly going on among the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand." In a similar vein, we suggest that the observed systematicity at the grammatical category level and the arbitrariness at the individual word level thus may be accounted for in terms of the learning, or processing, advantage that they induce. There are various proposals for the advantage of arbitrariness in the vocabulary. Some of these concern advantages for communicative systems, enabling the development of abstract terms as well as the facility to talk about events and objects that are distant in space and time (Clark, 1998; Hockett, 1960). Yet these communicative advantages also need to be part of a language that can be passed on from generation to generation as easily and accurately as possible. From this perspective the key question is what advantages can form-meaning arbitrariness confer to learning?

One explanation, and the one we favor, is that arbitrariness allows information present in the environment to have a maximum impact on the individuation of the intended word-referent mapping. As an example, consider a farmyard. Under normal circumstances the words for commonly observed referents are phonologically rather distinct: the possibility of mishearing *sheep* when someone says *cow* is thus minimal. However, if there was a close correspondence between form and meaning then the possibility of confusing the word for *sheep* with the one for *cow* is increased (e.g., if the two animals were referred to as *feb* and *peb*, respectively). Perceptual similarity among word forms would thus be an impediment to learning, as demonstrated in a computational model testing learning of arbitrary and systematic mappings for pseudo-patterns (Gasser, 2004).

The importance of arbitrariness for maximising the use of information from the environment was highlighted indirectly by the attempts to generate artificial languages that had a systematic relationship between sound and meaning (for a review, see Eco, 1995). These efforts indicated that confusions between similar sounding words were more likely to happen in systematic compared to arbitrary languages. For example, Wilkins' attempt to create a perfectly systematic language suffered from confusions between closely related terms. As noted by Eco (1995), in his *Essay*, Wilkins miswrote *Gade* (barley) for the similar meaning *Gape* (tulip),.

From an evolutionary perspective, Corballis (2002) suggested that arbitrariness ensures that we are less likely to confuse concepts that may be critical to survival. If the language was systematic, “edible plants or berries could be confused with poisonous ones, and animals that attack could be confused with those that are benign” (p.187), thus arbitrariness and systematicity becomes a matter of life and death. Because words are not spoken in a vacuum, but with a constellation of additional cues to the intended referent available to the listener, maximising the distinctiveness of the sound of the word itself facilitates perception, and therefore also learning, implying that homonymy can be tolerated when contexts are varied, so “bare” and “bear” will only co-occur in the same context on rare, and presumably rather unfortunate, occasions.

If the advantage of arbitrary form-meaning mapping derives from a need to minimise confusion between words with similar meaning, then semantically similar words have to occur often enough in similar contexts to cause confusion. Natural language does indeed seem to have the property that words’ meanings are encapsulated, or at least reflected, by the contexts in which they occur, which has been the foundation of some theories of meaning (e.g., Landauer & Dumais, 1997; Wittgenstein, 1953). However, exploiting this combination of information requires the listener to process multiple cues about the word’s identity. Returning to the farmyard example, *pig* is not generally confused with *wig* because these two words seldom occur in the same context (outside of Dr Seuss’ canon at least). To use knowledge of such differences and avoid confusing *wig* for *pig*, contextual information about the farmyard first needs to be integrated with the word form *pig*. Confusion is thus minimised when integration of information from numerous sources in the environment is maximised, via a vocabulary that has distinct sounds for words with similar meaning. Our experiments below demonstrate how such an integration of multiple sources can occur in language learning.

However, arbitrariness of the vocabulary, for all the advantages it confers, also comes with a cost in terms of impeding learning of similarities among words, as learning similarities among items is facilitated by shared phonological material (Brooks et al., 1993; Monaghan et al., 2005). Determining and processing these similarities is a task as critical for language learning as is individuating the words. As previously mentioned, we hypothesise that the structure of the vocabulary is poised between these two requirements: to individuate meanings and also to categorize over general semantic features. Hence, we propose that natural language incorporates a division of labor between systematicity and arbitrariness in the structure of the vocabulary. The role of morphology and other non-morphological prosodic and phonological indicators of grammatical categories in language, as well as the support for other clusters of words, such as expressives, indicates that systematicity can occur at a gross category level. However, the vocabulary also individuates each word as far as possible, given the category level constraints on the word's form. Our computational and experimental studies examine this division of labor in the vocabulary.

The first study presents a computational model trained on either arbitrary or systematic form-meaning mappings. We assess performance in terms of the two different tasks of key importance to early language acquisition: learning the overall category of different groups of words, and learning to individuate the precise referent for each word. Using a standard word to referent mapping task, the first experiment then tests predictions generated by the model regarding potential differences in the time-course and type of learning supported by arbitrary and systematic mappings. We hypothesise that a systematic relationship between phonology and categories will facilitate category learning. If our hypothesis that context must be integrated with the word's sound in order to observe an advantage for arbitrary mappings in learning is correct, such systematicity may also lead to better meaning individuation. However, if exploiting the distinct phonology of the arbitrary mappings does not depend on contextual information, contrary to our hypothesis, then we expect that an arbitrary relationship between phonology and categories will not be disadvantageous for learning the individual meanings of words.

The second simulation, in tandem with the second experiment, tests language learning when contextual information is also present in the learning environment. We hypothesise that the integration of multiple cues together with an arbitrary relationship between form and meaning will maximise the usefulness of information in the learner's environment and boost learning. With additional contextual information about the word, we predict that the systematic advantage for category learning will be reduced, and in addition, the arbitrary advantage will be observed for the task of individual word learning.

Simulation 1: Modeling systematic and arbitrary mappings without context

The first simulation investigated whether similarities between phonological forms with respect to their categories provide for better learning of the category membership of words, and the extent to which systematicity or arbitrariness, in the absence of contextual information, facilitated learning of individual word referents. We used connectionist models of learning, trained with backpropagation, to determine the ease at which phonological input patterns could be mapped onto output patterns representing two categories on the basis of general learning mechanisms.

Method

Networks. The architecture was a feedforward connectionist model, with 33 input units, 20 hidden units, and 10 output units. The input was constructed from three phoneme slots, representing each of the phonemes in a CVC monosyllabic word. Each phoneme was represented in terms of 11 phoneme features. The input layer was fully connected to the hidden layer, which was fully connected to the output layer. The number of hidden units was selected as the minimum number that resulted in reliable learning of the task.

Materials. The model was trained to map between input and output representations for 12 patterns. Each input pattern was instantiated as a CVC syllable so as to allow for the use of the same phonological forms in our human experiments. There were two sets of words with distinct phonological features. The first set had a fricative (/ʒ/ or /f/) in both onset and coda position, and contained either the vowel /i/ or /ɪ/. The second set had a plosive (/k/ or /g/) in both onset and coda, and contained either the vowel /a:/ or /u:/. The words used are shown in Table 1. The input pattern for each word was generated from the three phonemes of the words. Each phoneme was represented in terms of 11 phoneme features, taken from Harm and Seidenberg (1999), such that similar sounding phonemes had similar phoneme feature representations.

The output patterns were generated from two category prototypes. The first category was centred at values of .25 for each of the 10 output units, and the second category was centred at values of .75 for all 10 output units. Six individual output patterns were created for each category by randomly changing the values of each of the output units in the range +/-.25. There was, therefore, similarity between the output representations for patterns of the same category, and these were distinct from the output representations for the other category. Novel category patterns were randomly created for each simulation run.

Procedure. We trained the model on two conditions which held constant the set of input and output patterns, but altered the mapping between them. For the systematic condition, the patterns with similar phonological form were mapped onto output patterns of the same category. For the arbitrary condition, three of the fricative/front vowel words and three of the plosive/back vowel words were mapped onto each of the categories. To test that the phonological feature representation respected the similarities among the fricative/front vowel words and the plosive/back vowel words, we computed the Euclidean distance between all patterns within the same group of words, and all patterns across the two groups. The mean Euclidean distance within the groups was 2.91, SD = 1.09, and the mean distance across the groups was 4.72, SD = .53, which was significantly greater, $F(1, 64) = 77.32$, MSe= 53.55, $p < .001$.

Weights on connections between units in the model were initially randomised with mean 0 and uniform distribution in the range [-.5,.5]. The model was trained using backpropagation, such that after each pattern had been presented the weights on connections were adjusted to reduce the difference between the model's actual output and the target output representation for that pattern. We used a learning rate of .05, but learning rates of .1 and .01 resulted in similar performance, though with slightly quicker or slower learning, respectively. A training block consisted of the presentation of all 12 patterns in random order. Model performance was tested after 10, 20, 30, and 40 blocks of training.

During testing, we computed the Euclidean distance between the model’s actual activation and the target representation for each of the 12 patterns. For the categorization test, if the model’s actual activation was closest to a pattern of the same category, the model was judged to have learned the category for that item. If activation was closer to a pattern of the other category then a category error was recorded. For the individuation test, the model’s actual activation for a pattern had to be closest to the target representation for that pattern to be judged correct.

In addition to measuring the model’s performance after 10, 20, 30 and 40 blocks of training trials, in order to test the generality of these results we also recorded the number of training blocks needed before the model was able to reproduce (1) the correct category, and (2) the correct individuation response for all patterns. Longer training indicated slower learning.

The simulation was run 16 times, with different output category patterns, different random starting weights, and different random order of selection of patterns.

Results and discussion

The results for the category learning task for each of the four testing blocks are shown in Figure 1A. Accuracy of category responses was subjected to an ANOVA with Block as within-subjects factor and arbitrary/systematic Condition as between-subjects factor. As predicted, there was a main effect of Block[1], $F(3, 45) = 4.03$, $p < .05$, $\eta_p^2 = .21$, and there was a significant advantage for the systematic condition over the arbitrary condition, $F(1, 15) = 1228.84$, $p < .001$, $\eta_p^2 = .99$. This supported our prediction that learning the categories would be more accurate (and quicker) under the systematic than the arbitrary condition. The interaction between Block and arbitrary/systematic Condition was also significant, $F(3, 45) = 4.03$, $p < .05$, $\eta_p^2 = .21$, indicating that the additional training only affected the arbitrary condition, which showed overall lower accuracy and slower learning.

For the individuation task performance at the four testing blocks, the results are shown in Figure 1B. An ANOVA on accuracy of individuation responses indicated a significant main effect of Block, $F(3, 45) = 26.40$, $p < .001$, $\eta_p^2 = .64$, indicating that performance improved with time. There was a significant main effect of arbitrary/systematic Condition, $F(1, 15) = 39.19$, $p < .001$, $\eta_p^2 = .72$, with the systematic condition resulting in better learning. There was also a significant interaction between Block and Condition, $F(3, 45) = 5.45$, $p < .005$, $\eta_p^2 = .27$, indicating that the initial advantage for the systematic condition increased with more training. This is consistent with our hypothesis that integration with contextual information may be crucial to reveal a potential arbitrary advantage in language learning.

The amount of time taken to reach perfect performance is shown in Figure 2. For the categorization task, there was a significant effect of arbitrary/systematic condition, $t(15) = 22.12$, $p < .001$, with the systematic condition resulting in almost instantaneous

[1] In all cases where the main effect of block was significant, this was also significant in a linear contrast, with a larger effect size, indicating that performance improved over time.

learning of the category structure (mean 2.69 blocks SD = 1.35), but with much slower learning in the arbitrary condition (132.38 blocks, SD = 23.19). The difference in learning time between the arbitrary (312.19 blocks, SD = 91.83) and systematic conditions (280.38 blocks, SD = 90.72) was not significant for the individuation task, $t(15) < 1$. Though there was an initial advantage in learning individuation for the systematic condition, as indicated in the first forty blocks (as seen in Figure 1B), this advantage did not extend to learning to perfect performance for the whole set of mappings. For comparisons across the tasks, the time taken to achieve perfect performance was longer for the individuation than the categorization task for both the systematic and the arbitrary conditions, $t(15) = 19.66$, $p < .001$, and $t(15) = 5.59$, $p < .001$, respectively.

The results of the simulation show that a model that learns associatively the correspondences between words and two categories of patterns is advantaged by a systematic relationship between the sounds of words and the category structure, when the task is to learn the category structure. This is not surprising, given that feedforward connectionist models are able to exploit systematicity in mappings, and tend to find arbitrary input-output relations harder to learn and requiring of more resources (Lambon Ralph & Ehsan, 2006; Seidenberg & McClelland, 1989). It is also an explanation for why there appears to be systematicity at the category structure level in language (e.g., Kelly, 1992).

However, the model's performance in the early stages of learning was also significantly more accurate for the systematic compared to the arbitrary conditions for the individuation task: $p < .05$, for each of blocks 10, 20, 30, and 40. Still, the time taken to learn to perfect performance was not significantly different for the two conditions. The results thus point to a possible advantage for systematic mappings when learning takes place in the absence of any contextual information. Because the categories in the systematic condition were learned very quickly and before the first individuation test, the model could use such category information to reduce the problem of determining the specific referent for a word to one of six patterns. This provides an initial advantage over the arbitrary condition where the category structure is learned much more slowly and the model must select from one of twelve patterns, though eventually this systematic advantage reduces. We test this in the next set of simulations.

Examining the initial systematic advantage in the model

To test for a systematic advantage in the absence of contextual information, we conducted two additional simulations[2]. If the systematic advantage in the individuation task was because the systematic simulation could select from 6 rather than 12 patterns, then training the model with just 6 systematic patterns in a single output category should result in identical performance to the systematic condition, and be significantly better than the arbitrary condition. Furthermore, if our interpretation is correct then a simulation with 12 systematic patterns in a single output category should result in identical performance to the arbitrary condition for the individuation task, but worse performance than the original systematic condition.

For the simulation with 6 systematic patterns in a single category, the model was identical to that of the original systematic condition except that words from only one of

---

[2] We thank Amy Perfors for suggesting these simulations.

the categories were used during training and testing. The results of this new simulation were similar to the original systematic condition (.32 and .31 correct, respectively) but distinct from the original arbitrary condition (.18 correct). An ANOVA compared the new simulation to the systematic condition, with Block as an additional factor (10, 20, 30, and 40 blocks of training), and resulted in a significant main effect of Block, $F(3, 45) = 20.10$, $p < .001$, $\eta_p^2 = .60$, but no significant effect of Condition, or interaction, both $F < 1.23$. A second ANOVA comparing the new simulation to the original arbitrary condition with Block also as a factor resulted in a significant main effect of Block, $F(3, 45) = 12.17$, $p < .001$, $\eta_p^2 = .45$, a significant main effect of Condition, $F(1, 15) = 29.08$, $p < .001$, $\eta_p^2 = .66$, but no significant interaction, $F(3, 45) = 2.15$, $p = .11$, $\eta_p^2 = .13$.

For the simulation with 12 systematic patterns in a single category, we extended the set of patterns used in the original systematic condition to include all CVC words composed of all combinations of /ʒ/, /f/, and /v/ in onset and coda, and vowels /i/, and /I/. From these 18 possible words, we randomly selected 12 words for each simulation run, each of which mapped onto one of the categories in the model's output. The simulation was identical to the original one in every other way. The accuracy of the new simulation was similar to the original arbitrary condition (both .18 correct) but lower than the original systematic condition (.31 correct). An ANOVA on accuracy comparing the new simulation and the original systematic condition along with Block, resulted in a main effect of Block, $F(3, 45) = 33.86$, $p < .001$, $\eta_p^2 = .69$, a main effect of Condition, $F(1, 15) = 69.39$, $p < .001$, $\eta_p^2 = .82$, and a significant interaction, $F(3, 45) = 4.43$, $p < .01$, $\eta_p^2 = .23$. An ANOVA comparing the new simulation to the original arbitrary condition resulted in a significant main effect of Block, $F(3, 45) = 17.71$, $p < .001$, $\eta_p^2 = .54$, but no main effect of Condition and no interaction, both $F < 1.2$.

The results of these two additional simulations are consistent with our interpretation that the advantage of the original systematic condition over the arbitrary condition for the individuation task was due to the advantage in selecting from 6 items for responding in the systematic condition, and selecting from 12 items in the arbitrary condition.

Contrary to our simulations, Gasser (2004) found evidence for an arbitrary advantage when a large number of patterns in a dense representational space had to be learned. In Gasser's study, patterns varied over three dimensions and a systematic relationship was an advantage for learning an individuation task for 15 patterns, but not for 100 patterns. When the number of dimensions was increased to 4, however, the systematic advantage was found even for 100 patterns in his simulations. Natural language phonology permits many more dimensions than 4 for forming distinctions (in our simulations using simple CVC syllables there were, for instance, 11 features x 3 phonemes providing 33 dimensions), which is perhaps one reason why we did not replicate Gasser's (2004) findings in our simulations – the phonological space is too sparsely populated by the words in the language. As natural languages also exhibit sparsity in the sound space, it seems unlikely that Gasser's explanation for the arbitrariness advantage, as dependent on a densely populated representational space, can apply to natural language. Thus, it is critical to provide an empirical test of the model's predictions for a systematic advantage for both categorization and individuation when human learners are asked to learn the same materials.

Experiment 1: Learning systematic and arbitrary mappings without context

We tested participants' ability to learn mappings between the same CVC syllables used in the simulation studies and, as referents, pictures belonging to two distinct categories. Participants' learning was assessed in similar ways to the types of learning demonstrated in the models – we scored responses for each word in terms of learning categorization – whether the chosen referent was in the same category as the target referent – and in terms of learning individuation, when the target referent was correctly selected.

Method

Participants. Participants were 25 students at Queen's University in Canada, who participated for payment or course credit. One participant's data was not recorded due to a technical fault. All were native English speakers and had normal hearing and normal or corrected-to-normal vision. Of the 24 participants, 3 were male and 21 female, and mean age was 19.1 (SD = 1.5). Participants were randomly assigned to either the arbitrary or the systematic condition, with 12 participants in each condition.

Materials. For the words, the stimuli were the same 12 CVC syllables used in Simulation 1 (see Table 1). Each word was converted to speech with the FESTIVAL speech synthesizer (Black, Clark, Richmond, King, & Zen, 1994) using an adult male diphone voice set. All words were produced in monotone at 120Hz.

For the word meanings, we selected 32 black-and-white images from the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). To form two categories, we used six action images and six object images. For each participant, 6 action pictures and 6 object pictures were selected randomly from a set of 32 possible pictures (16 actions and 16 objects). The Appendix shows the list of picture topics. Each action picture illustrated a person performing some activity, and each object picture indicated only the object itself. Thus, the pictures formed quite distinct categories of actions and objects.

For the systematic condition, there was a correspondence between the grammatical category of the picture and the phonological features of the words, such that all the fricative words corresponded to the action pictures and the plosive words corresponded to the object pictures, or all the fricative words corresponded to the object and plosives to the action pictures. Whether the fricatives matched the action or the object pictures was counterbalanced across participants in order to control for possible preferences between certain words and categories. For the arbitrary condition, half the fricative words and half the plosive words corresponded to the action pictures, and the remaining words corresponded to the object pictures. For the arbitrary condition, the word grouping was again counterbalanced for which group of words related to the action and object pictures.

Procedure. Participants were instructed that they were to learn an alien language, in which they would hear a sentence in the language and see a picture. They were further asked to learn the relationship between the sentence and the picture. Participants were trained on the word-picture pairings and then tested on their ability to match each word to the pictures. This sequence of training and testing was repeated four times.

For each training trial, the picture was presented at the centre of a computer screen and simultaneously the word was played through headphones. After 2000ms from onset of the word, the next trial began. During a training block, each picture and word was presented four times each, making for a total of 48 learning trials in each block.

For testing, all 12 pictures were presented simultaneously on the screen, in three rows of 4 pictures. Each picture was labeled with a number from 1-12, and the order of the pictures was randomized for each participant and for each testing block. The pictures were presented for 5000ms, and then the testing trials began. For each testing trial, one of the words was played over the headphones, and the participant had to indicate with a key press on a keyboard labeled with the numbers 1-12, which picture corresponded to the word. When a response was made, the next word was played, and no feedback was given. There were 12 trials, one for each word. As for the Simulation studies, we had two dependent measures. To assess category learning, we tested the number of responses that were to words of the same category (object or action). To assess individuation of word meaning, we tested the number of responses that correctly identified the precise referent of the word.

Results and discussion

The results in terms of same-category responses in the arbitrary and systematic conditions are shown in Figure 3A. Participants were able to solve the categorization task. For both conditions, at all testing points, performance was significantly more accurate than chance level of .5, $p < .001$, except for block 1 in the arbitrary condition, $p = .54$. We conducted an ANOVA on accuracy of category responses, with testing Block (1- 4) as a within-subjects factor, and Condition (arbitrary/systematic) as a between-subjects factor. There was a main effect of Block, $F(3, 66) = 11.38$, $p < .001$, $\eta_p^2 = .34$, indicating that participants improved with additional exposure on the task. There was also a main effect of Condition, $F(1, 22) = 140.25$, $p < .001$, $\eta_p^2 = .86$, with an advantage for the systematic condition over the arbitrary condition in terms of accuracy. This corresponds to the quicker learning of the category structure in the systematic condition in Simulation 1 (compare Figures 1A and 3A). The interaction was not significant, $F(3, 66) = 1.65$, $p = .19$, $\eta_p^2 = .07$.

For the individuation task, the results are shown in Figure 3B. Participants were able to solve the task in both conditions. All testing blocks for both conditions were significantly above chance level of .08 correct, $p < .001$, apart from the first testing block in the arbitrary condition, which was marginally significant (with Bonferroni correction, $p = .07$). A repeated measures ANOVA with Block as a within-subject variable and arbitrary/systematic Condition as a between subject variable, revealed a significant main effect of Block, $F(3, 66) = 17.30$, $p < .001$, $\eta_p^2 = .44$, again showing that performance improved with time. However, there was no main effect of Condition, $F(1, 22) = 2.77$, $p = .11$, $\eta_p^2 = .11$, which differed from the Simulation 1 results for the first forty blocks of training by the model (Figure 1B), though reflected the simulation results for extended training (Figure 2). In the experimental results, early in training in block 1 there was a marginally significant advantage for the systematic condition, $p < .10$ (at all other testing blocks, $p > .36$). As for the same-category responses, there was no significant interaction between Block and Condition, $F < 1$.

The behavior in the learning study resembled the main pattern of results for the model in Simulation 1. Systematic mappings resulted in more accurate learning of the category distinction between the action and object labels – both the model and the behavioral data demonstrated a main effect of condition for the category response, and the model learned to categorize all the patterns at an earlier stage in the systematic condition. Furthermore, both Simulation 1 and Experiment 1 indicated a smaller difference between the conditions for the individuation task: there was no main effect of Condition in the behavioral study, and the model's learning to perfect accuracy took similarly long in both conditions.

So far we have explored the role of systematicity and arbitrariness in a task resembling standard word learning paradigms (e.g., Xu & Tenenbaum, 2007; Yoshida, Fennell, Swingley, & Werker, 2009), and found no advantage for arbitrariness in the mappings for the simulation nor for the experimental study. However, as we have earlier hypothesised that the presence of contextual information may affect learning substantially, and indeed that a confluence of cues may be critical for exhibiting the arbitrary advantage, we next present a computational instantiation of this hypothesis. Specifically, we measure the ability of a computational model to exploit the arbitrary mapping for learning to individuate patterns when additional contextual information is provided to the model. We predicted that adding the contextual cue would facilitate learning to individuate meanings, particularly in the case of the arbitrary mappings.

Simulation 2: Simulating systematic and arbitrary mappings with context

Simulation 2 was identical to Simulation 1, except that additional contextual information indicating the category of the word was also provided as input to the model. Contextual information may result from one of many co-occurring sources, such as gesture, eye-gaze, or environmental saliency (e.g., Brandone, Golinkoff, Pence, & Hirsh-Pasek, 2007; Gliga & Csibra, 2009; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Tomasello, 2003). But contextual information can also arise from within the language itself (Monaghan et al., 2007). For instance, in child-directed speech, nouns and verbs tend to be reliably preceded by distinct highly-frequent words (Monaghan et al., 2005), and such language-internal contextual information can provide useful constraints on the potential meaning of the word, as in the contextual co-occurrence models of meaning (e.g., Landauer & Dumais, 1997). In Simulation 2, this contextual information was operationalised as additional units that were perfectly predictive of the category to which the word belonged, and so can be seen as either a morphological marker indicative of grammatical category, or a function word, such as "the" or "you" which would reliably predict the grammatical category of the following word ("the" tends to precede nouns, and "you" tends to precede verbs). Such information has been shown to be available in natural language (St. Clair, Monaghan & Christiansen, 2010) and useable for categorization in artificial language learning studies (Mintz, 2002; St. Clair et al., 2009). The context in the model, and in the behavioural experiments to follow, refers to information that is available to the learners in addition to the form-meaning pairings, the latter of which is typically the sole source of information in word learning studies (as reflected in Simulation 1 and Experiment 1).

Method

Network. The model's architecture was identical to that of Simulation 1, except that an additional context input layer was added to the model, fully connected to the hidden layer. The input layer comprised two units, one representing each of two function words in the language.

Materials. The input and output patterns were the same as for Simulation 1, though with additional units in the model's input where contextual information was provided.

Procedure. Training and testing was conducted in the same way as in Simulation 1, except that for every word input pattern presented to the model, the relevant context input layer unit was also activated. This unit was perfectly predictive of the category to which the word belonged. Performance was assessed as in Simulation 1: when the model produced an output pattern closest to a pattern of the same category as the target, then it was judged to have learned the category of the word; when the model produced an output pattern closest to the target pattern, the model was judged to have learned to individuate that particular word. We again measured the performance of the model on the same-category response and the individuation task after ten, twenty, thirty and forty blocks of training. We also trained the model until its performance on each task reached 100% accuracy. 16 runs of the model were conducted, with different random starting weights, and different random permutations of the patterns.

Results and discussion

The results for the model after 10, 20, 30, and 40 blocks of training for the category response and for the individuation task are shown in Figure 4. For the categorization task, an ANOVA with the four Blocks of training as within-subjects factor and arbitrary/systematic Condition as between-subjects factor revealed a main effect of Block, $F(3, 45) = 30.89$, $p < .001$, $\eta_p^2 = .67$, showing that performance improved with increased training. There was also a main effect of arbitrary/systematic Condition, $F(1, 15) = 42.64$, $p < .001$, $\eta_p^2 = .74$, with more accurate category responses for the systematic condition. There was also a significant interaction, $F(3, 45) = 30.89$, $p < .001$, $\eta_p^2 = .67$, due to the ceiling performance of the systematic condition and the improving performance with time for the arbitrary condition (the same *F*-value as for the main effect of Condition was due to perfect performance in the systematic model). In contrast to Simulation 1, which had no extra context information, the arbitrary condition in Simulation 2 reached ceiling performance of 100% accuracy before the end of 40 blocks of training. The model thus utilised the contextual information to guide categorization performance, though, in the arbitrary condition, this was not achieved immediately but after a short training period.

For the individuation task, an ANOVA on accuracy of responses with Block and arbitrary/systematic Condition as factors resulted in a main effect of Block, $F(3, 45) = 51.86$, $p < .001$, $\eta_p^2 = .86$, but, unlike for Simulation 1, the main effect of arbitrary/systematic Condition was not significant, $F(1, 15) = 2.14$, $p = .16$, $\eta_p^2 = .12$. The interaction, however, was significant, $F(3, 45) = 3.87$, $p < .05$, $\eta_p^2 = .21$. There was an initial advantage in training for the systematic condition (block 10: $p < .05$), due to the model being able to select within the category for this condition. However, critically, later in training there was an advantage for the arbitrary condition (block 20: *ns*; block 30: $p < .05$; block 40: $p = .08$). The increase in accuracy for the arbitrary condition (mean = .34, SD = .15) was greater than for the systematic condition (mean = .21, SD = .13), $t(15) = 2.18$, $p < .05$.[3]

For the number of blocks taken to achieve perfect performance on the same-category response, the systematic condition (mean = 3.00, SD = 1.59) was, as predicted, significantly quicker than the arbitrary condition (mean = 17.94, SD = 5.23), $t(15) = 12.72$, $p < .001$. Even though the context layer provided perfect information about the category structure, the conformance of the phonological representation to the category structure still resulted in quicker learning, as shown in Figure 5. However, the difference between the systematic and arbitrary condition in learning time was greatly reduced compared to Simulation 1. In Simulation 1, the mean difference in learning time was 129.69 blocks of training, whereas for Simulation 2 it was 14.94 blocks, which was significantly less, $t(15) = 18.31$, $p < .001$.

For the time taken to achieve perfect performance on the individuation task, the arbitrary condition (mean = 190.44 blocks, SD = 71.45) resulted in quicker learning of the whole word set than the systematic condition (mean = 277.56, SD = 74.06), $t(15) = 3.85$, $p < .005$ (see Figure 5). For comparisons with the categorization task, learning was slower for the individuation task for both the systematic and arbitrary conditions, $t(15) = 22.61$, $p < .001$, and $t(15) = 11.16$, $p < .001$, respectively.

---

[3] We used pairwise comparisons for the model results, as each of the systematic and arbitrary simulation runs can be directly compared as they were controlled to have the same initial random starting weights.

Simulation 2 supported our hypothesis about the importance of combining arbitrariness with contextual information in order to reveal the learning advantage for arbitrary mappings. The simulation showed that, if the context provides information about the general meaning of the word during learning, then there is a small initial advantage in the first blocks of training for the systematic condition, because the selection can be restricted to within words of the same category – therefore the model is selecting amongst 6 words rather than 12. However, with additional training, this initial disadvantage for the arbitrary condition reduces, until, as shown in Figure 4B, an advantage for arbitrary mappings between the sounds of words and their meanings emerges. This was due to the greater variance in the phonology of words within each category, which could provide more opportunities for distinguishing words within those categories. When the model was trained to perfect performance, the model trained on the arbitrary mappings learned more quickly than when trained on the systematic mappings. If information can be integrated between contextual information and the sound of the word, then this maximising of information is beneficial for learning, and offers a potential explanation for how arbitrariness in the vocabulary is not only tolerated but may be critically advantageous for language acquisition.

The next experiment tested whether, as predicted by the model, participants would also be able to utilise contextual information to assist in forming sound-meaning mappings for words. Providing contextual cues for word-learning would, we predicted, result in an advantage for arbitrary mappings, which was not found, in Simulation 1 and Experiment 1, when no contextual information was present.

Experiment 2: Learning systematic and arbitrary mappings with context

This experiment was identical to that of Experiment 1, except that additional contextual information was provided to delineate the categories of the words. As in Experiment 1, we compared learning relations between the same set of words and the same set of pictures varying whether the relationship between sounds of words and the category to which the picture belonged was systematic or arbitrary. Distinct from Experiment 1, we also added a contextual cue in the form of a preceding word that reliably denoted each category of words.

Method

Participants. 24 participants from Queen's University took part in the study. All had normal hearing and normal or corrected-to-normal vision. None had participated in Experiment 1. 7 were male, 17 female, with mean age 19.4 years (SD = 1.6). There were 12 participants in each of the arbitrary and systematic conditions.

Materials. The materials were identical to those used in Experiment 1, except that there were two additional context words. Using the same speech synthesiser and diphone database, we generated two CV syllables: /wɛ/ and /mə/.

Procedure. The procedure was identical to Experiment 1, except that participants were provided with contextual information about the category of the word. In this experiment, contextual information was presented within the speech stream (Monaghan & Mattock, 2009), though we predict that other sources of contextual information are also likely to affect learning in similar ways, when they provide constraints on the referent's category. Participants heard each referring word preceded by a context word during the training and testing phases. All the words in one group were always preceded by one of the context words, and all the words in the other group were always preceded by the other context word. The context words therefore provided reliable information about the category of the picture, i.e., whether it was an object or an action. The assignment of the context word to each group was randomised across participants, in order to avoid any latent preferences for the context word relating to either actions or objects.

For the training trials, the context word was played simultaneously with the picture appearing, then after 300ms the referring word was played. After 1700ms from onset of the referring word, the next trial began. For the testing trials, the context word and the referring word were played again with a 300ms interval.

Results and discussion

For the proportion of same-category responses, the results are shown in Figure 6A. As for Experiment 1, participants were able to learn to respond with the same-category greater than chance level of .5 in all conditions for all blocks, $p < .005$, except for the arbitrary condition in block 1, $p = .10$. An ANOVA with Block and arbitrary/systematic Condition as factors revealed a significant main effect of Block, $F(3, 66) = 12.99$, $p < .001$, $\eta_p^2 = .37$, showing that participants' learning improved with exposure. As with Simulation 2, there was a significant main effect of Condition, $F(1, 22) = 48.90$, $p < .001$, $\eta_p^2 = .69$, with the systematic condition resulting in significantly more accurate category learning than the arbitrary condition. There was also a significant interaction, $F(3, 66) = 10.17$, p

$< .001$, $\eta_p^2 = .32$, due to the initial significant difference for same-category responses between the systematic and arbitrary conditions in Blocks 1, 2, and 3, $p < .001$, $p < .001$, and $p < .05$, respectively, which was reduced in block 4 to a non-significant difference, $p = .16$. After sufficient exposure to the language, participants were able to use the additional contextual information to support their learning of the category structure of the task, even when the relationship between the sounds of words and the categories is arbitrary. This was indicated by the increased same-category responses for the arbitrary condition in Experiment 2 as compared to Experiment 1 (*t*-tests across the two experiments revealed significant difference in performance for Blocks 2, 3, and 4, all $p < .05$, but not for block 1, $p = .34$).

For the proportion of pictures correctly individuated, the results are shown in Figure 6B. As in Experiment 1, participants were able to solve the task. For all testing blocks for both conditions performance was significantly better than chance level of .08, $p < .001$, except for the first testing block for the arbitrary condition, $p = .06$. An ANOVA resulted in a significant main effect of Block, $F(3, 66) = 19.07$, $p < .001$, $\eta_p^2 = .46$, showing again that extended exposure resulted in better performance. As with Simulation 2, there was no significant main effect of Condition, $F < 1$. However, distinct from Experiment 1 and in line with the predictions from Simulation 2, there was a significant interaction between Block and Condition, $F(3, 66) = 3.93$, $p < .05$, $\eta_p^2 = .15$. This was because the accuracy of individuation responses in the arbitrary condition increased more rapidly than for the systematic condition, thus the interaction effect mirrored the results of Simulation 2. There was an initial slight disadvantage for the arbitrary condition because participants had to select the correct picture from a choice of

12 rather than a choice of 6 for the systematic condition. This was because in the arbitrary condition the sound of the word did not facilitate a selection just from words of the same category. The comparisons between Experiment 1 and 2 indicated that the context information was only effective in producing more same-category responses for blocks 2 through to 4. Participants in block 1 had not yet learned to map this context information onto the category distinction for the pictures. But later, when participants learned to integrate the contextual information with the sound of the word to determine the picture to which it referred, there was an emerging advantage for the arbitrary condition. As with the results of Simulation 2, the increase in proportion correct for the arbitrary condition between blocks 1 and 4 was .35 (SD = .16), which was significantly larger than the .18 (SD = .21) increase obtained for the systematic condition, $F(1, 22) = 4.73$, $p < .05$, $\eta_p^2 = .18$.

The results indicated that, after multiple exposures to the language, participants were able to integrate contextual information from the preceding context marker with the distinct sound of the target word. This integration resulted in a final categorization performance for the arbitrary condition at a level converging with that of the systematic condition, and achieving a level of accuracy higher than when no contextual information was present. When context was present there was a large category learning improvement compared to when no contextual information was present, which replicated the category learning effects of the simulation studies.

The integration of information from the marker-word context with the sound of the word also resulted in quicker improvement of learning the individual referents for words in the arbitrary condition. Simulation 2 predicted that the presence of contextual information, for a learning system responding to the correlated information between input and output patterns, should result in an arbitrary advantage. In this experimental study, we have shown that embedding the contextual information within the speech in the form of co-occurring words can result in the same arbitrary advantage, reflected in the significant interaction between block and arbitrary/systematic condition. However, insight from simulating the experimental results in computational models can be further enhanced by assessing the mechanisms by which the models solve the task. The next section provides further tests of our hypotheses about the contributions of contextual information to the learning of arbitrary versus systematic mappings.

Examining the arbitrary advantage

In order to provide insight into the way in which contextual information may promote the arbitrary advantage, we investigated the nature of the internal representations that the models developed during learning. Specifically, we analysed the pattern of hidden unit activations that the models produced when mapping from a particular input to a particular output. Probing these activation patterns thus allows us to go beyond mere assessments of output behavior to determining the kind of internal similarity structure that the models develop in order to accomplish the mapping task. We compared each model – arbitrary and systematic – both with and without context, examining their representations early in training (after 10 training epochs) and at the end of training (40 epochs). For each input pattern, the hidden unit activations were recorded and entered into a principal components analysis with two components extracted using varimax rotation. The resulting loadings on the first two components for each pattern are shown in Figure 7 for one simulation run of each model. The first two components accounted for 54.5% of the variance for the arbitrary condition with no context, 56.7% for the arbitrary condition with context, 81.2% of the variance for the systematic condition with no context, and 72.4% for the systematic condition with context.

After 10 epochs of training, for the systematic condition both with and without context information demonstrate that the category structure has been learned effectively, and is reflected along the first component. For the arbitrary condition with context, the category structure is emerging, but with a degree of overlap along the first component. For the arbitrary condition with no context the category structure is not reflected at all. After 40 epochs of training, the systematic conditions continue to represent the category structure effectively. The arbitrary condition with context has also learned the category structure effectively, but this is not reflected in the arbitrary condition with no context. The principal components analysis also provides clues as to why individuation was not poorer in the systematic condition without context (from Simulation 1 and Experiment 1). In this condition, though the model's hidden unit representations appear to be not so distinct as for the arbitrary condition, there is still sufficient distinction between each of the representations to support individuation. If the phonological space was much more dense, then the arbitrary advantage for individuation may be observed even without contextual information due to yet poorer distinctions between items for the systematic condition (see Gasser, 2004, for a related argument).

To determine quantitatively the extent to which the category structure is reflected in the hidden unit representations over time, we measured the Euclidean distance between hidden unit representations for every pair of patterns. We then determined the mean distance between pairs of patterns of the same category, and the mean distance between pairs of patterns of different categories. We conducted an ANOVA on Euclidean distance with Condition, Context, whether the distance was measured Within/Between categories, and training Block as within subject variables. The results are shown in Figure 8. All main effects and interactions were significant, so we concentrate on only those effects that relate to the hypotheses about the representational bases of the models' performance. The greater integrity of the category representations in the systematic conditions compared to the arbitrary conditions was reflected in the significant interaction between Condition and Within/Between, $F(1, 15) = 11521.44$, $p < .001$, $\eta_p^2 = 1.00$. This was modulated by the effect of Context, as the three-way interaction was also significant, $F(1, 15) = 450.68$, $p < .001$, $\eta_p^2 = .97$. The four-way interaction with Block was also significant, $F(1, 15) = 1509.39$, $p < .001$, $\eta_p^2 = .99$, which is shown in Figure 8 by the increasing difference of Within/Between after 40 epochs compared to 10 epochs, particularly for the arbitrary condition with context. Hence, this four-way interaction indicates that category structure is represented earlier in the systematic conditions, but that the arbitrary condition with contextual information permits a gradual learning of the category structure.

It is possible that the early representation of category structure may have inhibited effective learning for the individuation task for the systematic conditions[4]. The Principal Components analyses shown in Figure 7 suggest that the representations in the systematic conditions may have been less distinct. To test this in the model, we determined for each

---

[4] We thank an anonymous reviewer for pointing out this potential cause, and for suggesting some of the analyses of the model's representations included in this section.

representation the smallest distance to the representation of another pattern. Large distances indicate that the model's representation of the pairings is distinct, smaller distances indicate less effective separation of the representations. The results are shown in Figure 9. We conducted an ANOVA on mean minimum distance between patterns, with Condition, Context, and training Block as within-subject variables. There was a significant main effect of Condition, $F(1, 15) = 846.94$, $p < .001$, $\eta_p^2 = .98$, with smaller distances for the systematic models, as predicted, indicating reduced individuation for these simulations. There was also a significant main effect of Context, $F(1, 15) = 98.15$, $p < .001$, $\eta_p^2 = .87$, with the presence of context increasing inter-item distances. The main effect of Block was also significant, $F(1, 15) = 174.46$, $p < .001$, $\eta_p^2 = .92$, with increasing distances with increased training. As shown in Figure 9, there was a significant interaction between Condition and Context, $F(1, 15) = 126.85$, $p < .001$, $\eta_p^2 = .89$, as context had a greater effect on the arbitrary condition. Also, the interaction between Condition and Block was significant, $F(1, 15) = 90.75$, $p < .001$, $\eta_p^2 = .86$, with greater increase in distances with time for the arbitrary condition. No other interactions were significant, Fs < 1.07.

Hence, in the model the advantage of arbitrary representations coupled with contextual information is twofold: it minimizes confusion between items within categories, and it permits the task of discovering category structure to be achieved without negatively affecting the learning of names of individual referents. This is because context for the arbitrary condition provides complementary information for word identification, whereas for the systematic condition context overlaps with information already present within the word's phonology.

In natural language learning situations, however, contextual information is not always available. For low-frequency words or words that are novel, for instance, the reliability of co-occurrence information in speech is low, and so reliance on the word's form only may be the best source of information about the word's meaning and correct usage (Monaghan et al., 2005). Consequently, a truly optimal solution for the vocabulary may be to incorporate both unique, discriminating information as well as shared, category-level information within the phonological form of the word. In the next simulation we test whether this combination can serve both the individuation as well as the categorization tasks, without a substantial cost in learning for either task.

Simulation 3: Balancing individuation and categorization in the mappings

This simulation extended the first simulation study by adding an additional condition such that half the phonological representation of the word was arbitrarily related to the output representation, and the other half of the phonological representation was correlated with the output. We refer to this as the halfhalf condition. We hypothesized that this representational format would allow the model to take advantage of the most useful properties of both arbitrary and systematic form-meaning mappings in regard to learning. In particular, we predicted that the model would be able to capitalize on systematicity for category learning and arbitrariness for meaning individuation—even in the absence of additional contextual information. We also tested whether additional contextual information for the halfhalf condition altered the model's behaviour in any way.

Method

Network. For the halfhalf model with no context, the architecture was the same as for Simulation 1, and there was no context layer in the model. For the halfhalf model with context, the architecture was identical to Simulation 2, in that there was a context layer.

Materials. The model was trained to map between input and output representations for 12 patterns. The same set of phonemes used to construct the stimuli in Simulations 1 and 2 were used, but the relationship between the onset and the vowel of the word and the output category was arbitrary, and the relationship between the coda and the output category was systematic. The codas /ʒ/ or /f/ ended words in the first category, and the codas /k/ or /ɡ/ ended words in the second category. The vowels /i/, /ɪ/, /a:/ and /u:/ occurred in words of both categories, as did the onsets /ʒ/, /f/, /k/ and /ɡ/. The words are shown in Table 2. As in Simulations 1 and 2, the words were converted into phoneme feature representations, and the output category representations were constructed in the same way as for these other simulations. The context information was provided for the halfhalf with context model, just as for Simulation 2.

Procedure. The model was tested in the same way as for Simulations 1 and 2, assessing individuation of meaning as well as categorization. As for Simulations 1 and 2, the simulation was run 16 times, with different output category patterns, different random starting weights, and different random ordering of patterns.

Results and discussion

We compared the halfhalf model without context to the results of Simulation 1, and the halfhalf model with context to the results of Simulation 2. The critical test is the halfhalf model's performance against the results of Simulation 1 – good performance of the halfhalf model without context would indicate that contextual information was established sufficiently within the word itself. We also analysed the halfhalf model with context to establish whether the results of the model are stable, or whether the additional contextual information interferes in the model's processing of the contextual information within the word.

For the halfhalf model without context, results for categorization and for individuation are shown in Figure 10, in comparison with the results of Simulation 1. We conducted planned comparisons to test whether learning of the halfhalf mappings without context was similar in performance for categorization to the systematic mapping without context, and distinctly better than the arbitrary mapping without context. An ANOVA with Block and Condition (systematic versus halfhalf) on categorization performance indicated a significant main effect of Condition, $F(1, 15) = 20.72$, $p < .001$, $\eta_p^2 = .58$, and a significant main effect of Block and a significant interaction, both $F(3, 45) = 20.03$, $p < .001$, $\eta_p^2 = .57$. The halfhalf condition resulted in slightly slower learning to 100% accuracy than the systematic condition, resulting in the main effect and the interaction. However, by 40 training blocks, the performance of the halfhalf condition was identical to the systematic condition, resulting in 100% accuracy.

The ANOVA comparing the arbitrary condition without context to the halfhalf condition without context for categorization resulted in a significant main effect of Block, $F(3, 45) = 15.37$, $p < .001$, $\eta_p^2 = .51$, a significant main effect of Condition, $F(1, 15) = 750.75$, $p < .001$, $\eta_p^2 = .98$, but no significant interaction, $F(3, 45) = 1.60$, $p = .20$, $\eta_p^2 = .10$. The halfhalf condition was significantly more accurate than the arbitrary condition for categorizing the patterns when no context was present, and this effect pertained at all testing blocks, all $p < .01$.

For the individuation task, for the planned comparison between the systematic condition without context and the halfhalf condition without context, there was a significant main effect of Block, $F(3, 45) = 51.61$, $p < .001$, $\eta_p^2 = .77$, a significant main effect of Condition, $F(1, 15) = 6.79$, $p < .05$, $\eta_p^2 = .31$, but no significant interaction, $F < 1$. The halfhalf condition resulted in more accurate individuation than the systematic condition. Initially and in latter stages of training there was no difference in performance between the systematic and the halfhalf conditions, $p >= .15$, but there was a significant difference for the 20th block of training with greater accuracy for the halfhalf condition, $p < .05$. For the comparison between the halfhalf mapping and the arbitrary mapping for individuation when context was not present, there was a significant main effect of Block, $F(3, 45) = 28.75$, $p < .001$, $\eta_p^2 = .66$, a significant main effect of Condition, $F(1, 15) = 73.92$, $p < .001$, $\eta_p^2 = .83$, and a significant interaction, $F(3, 45) = 10.71$, $p < .001$, $\eta_p^2 = .42$. The main effect of Condition was due to the higher accuracy of the halfhalf mapping than the arbitrary mapping overall, .37 and .18, respectively, which was a difference that increased with training.

The results of the halfhalf condition with context are shown in Figure 11, in comparison with the results of Simulation 2. For the categorization task, in a planned comparison between the halfhalf condition with context and the systematic condition with context, as the only difference was for the first training Block (halfhalf model resulted in 99.0% correct, compared to 100% correct for the systematic condition with context), the main effects of Block, Condition, and the interaction effect were identical F

= 7.50, $p < .05$, $\eta_p^2 = .33$. A planned comparison with the arbitrary condition with context resulted in a significant main effect of Block, $F(3, 45) = 31.59$, $p < .001$, $\eta_p^2 = .68$, a significant main effect of Condition, $F(1, 15) = 38.36$, $p < .001$, $\eta_p^2 = .72$, and a significant interaction, $F(3, 45) = 25.13$, $p < .001$, $\eta_p^2 = .63$. The arbitrary condition with context resulted in lower initial accuracy, but greater improvements in accuracy over time.

For the individuation task, the comparison of the halfhalf condition to the systematic condition with context resulted in a significant main effect of Block, $F(3, 45) = 32.77$, $p < .001$, $\eta_p^2 = .69$, a significant main effect of Condition, $F(1, 15) = 19.61$, $p < .001$, $\eta_p^2 = .57$, but no significant interaction, $F < 1$. The difference in accuracy was significant for block 10 and 30, both $p < .05$, but not for the other blocks, both $p > .10$. Comparing the halfhalf condition to the arbitrary condition with context, there was a main effect of Block, $F(3, 45) = 80.60$, $p < .001$, $\eta_p^2 = .84$, a marginally significant effect of Condition, $F(1, 15) = 4.25$, p = .06, $\eta_p^2 = .22$, and no significant interaction, $F(3, 45) = 2.12$, $p = .30$, $\eta_p^2 = .12$.

The next study tests experimentally the predictions of Simulation 3 that the halfhalf condition is an optimal compromise to serve both categorization and individuation.

Experiment 3: Testing the halfhalf language

We used exactly the same stimuli as were presented to the model in the halfhalf simulation. In the behavioural study, we compared the halfhalf condition to a fully arbitrary condition with no context in which there was no relationship between the word beginning or the word ending and the category of the word. This was the critical comparison – to determine whether including category information within the word, at the expense of fully arbitrary information, could improve both categorization and individuation performance over a context free arbitrary condition. We also compared the halfhalf condition against the arbitrary and systematic conditions without and with context from Experiments 1 and 2 to determine whether the halfhalf condition was an optimal balance of information for categorization and individuation, though the results of cross-experiment comparisons should of course be interpreted with caution.

For the simulation study, there should be little difference in performance for the arbitrary with context condition and the halfhalf condition without context, as in both these conditions there is systematic and arbitrary information supporting categorization and individuation, respectively, however the quantity of the information for individuating in the halfhalf simulation was somewhat reduced. For the behavioural study of the halfhalf condition, the quantity of the context and the quantity of the arbitrary information was reduced compared to the previous experiment, and also the ordering of the arbitrary and systematic information was exchanged. For the model, the information from all parts of the word arrived simultaneously at the hidden layer, so this made no difference. Hence, the experiment presents a robust test of the benefit of optimizing information within the word for categorization and individuation.

Method

Participants. 24 undergraduate students from Queen's University participated in the study for course credit or payment. Seven were male and 17 female, with mean age 20.3

years (SD = 1.7). No participant had taken part in any other study reported in this paper, all reported having English as a first language, and had normal or corrected to normal hearing and vision. Participants were randomly assigned to either the halfhalf or the arbitrary condition.

Materials. The same words as used for Simulation 3 were employed in the study, with sound produced by the FESTIVAL (Black et al., 1994) speech synthesizer in precisely the same way as for Experiments 1 and 2. For the halfhalf condition, the pictures were selected randomly from the same set as were used in Experiments 1 and 2, and were paired with the two categories of words, with different random pairings within the categories for each participant.

For the arbitrary condition, three of the words within each category were exchanged with three of the words from the other category. Hence, neither the sound of the beginning nor the end of the word corresponded to the action/object category distinction. The words to be exchanged were selected randomly for each participant.

Procedure. The procedure was identical to Experiments 1 and 2.

Results and Discussion

Learning for both categorization and individuation was effective in the halfhalf condition, with .81 and .61 words correctly classified, respectively. For the arbitrary condition (which was without context), performance was less accurate at .73 and .48 for categorization and individuation, respectively. The results across the four blocks of testing are shown in Figure 12. Figure 12 also shows the results of the systematic condition of Experiment 1, for comparison. An ANOVA with Block (1- 4) and Condition (Arbitrary/Halfhalf) as factors, on categorization accuracy resulted in a main effect of Block, $F(3, 66) = 10.66$, $p < .001$, $\eta_p^2 = .33$, with better performance with increasing training, a main effect of Condition, $F(1, 22) = 4.43$, $p < .05$, $\eta_p^2 = .17$, but no significant interaction, $F < 1$. An ANOVA with Block and Condition as factors on individuation performance resulted similarly in a main effect of Block, $F(3, 66) = 42.67$, $p < .001$, $\eta_p^2 = .66$, a main effect of Condition, $F(1, 22) = 5.11$, $p < .05$, $\eta_p^2 = .19$, and no significant interaction, $F < 1$. The halfhalf condition therefore resulted in better categorization and individuation performance than the arbitrary condition without context, consistent with the predictions of Simulation 3.

These results were confirmed in a planned comparisons against the arbitrary condition without context from Experiment 1, with Block and halfhalf/arbitrary Condition as factors. For the categorization task, there was a significant main effect of Block, $F(3, 66) = 9.36$, $p < .001$, $\eta_p^2 = .30$, a significant main effect of Condition, $F(1, 22) = 53.08$, $p < .001$, $\eta_p^2 = .71$, but no significant interaction, $F < 1$. For the individuation task, there was a significant main effect of Block, $F(3, 66) = 37.27$, $p < .001$, $\eta_p^2 = .63$, a significant main effect of Condition, $F(1, 22) = 23.57$, $p < .001$, $\eta_p^2 = .52$, but no significant interaction, $F = 1.27$. Hence, the halfhalf condition was more successful than the arbitrary condition without context for participants' learning both categories and individual meanings of words.

In planned comparisons against the arbitrary condition with context from Experiment 2, there was again a significant main effect of Block, $F(3, 66) = 14.63$, $p <$

.001, $\eta_p^2 = .40$, and of Condition, $F(1, 22) = 5.81$, $p < .05$, $\eta_p^2 = .21$, but no significant interaction, $F(3, 66) = 2.29$, $p = .09$, $\eta_p^2 = .09$, for categorization. For individuation, similarly, there was a main effect of Block, $F(3, 66) = 44.08$, $p < .001$, $\eta_p^2 = .67$, Condition, $F(1, 22) = 11.97$, $p < .001$, $\eta_p^2 = .35$, and no significant interaction, $F(3, 66) = 2.10$, $p = .11$, $\eta_p^2 = .09$. The pattern of results was similar to the halfhalf simulation. This suggests that the halfhalf condition was effective in conveying the division of labor in terms of systematicity supporting categorization and arbitrariness combining with context to support individuation. Indeed, incorporating the phonology at the end of the word was more effective in supporting categorization than having a separate context cue preceding the word, as in Experiment 2.

In a planned comparison between the halfhalf condition and the systematic condition without context from Experiment 1 (as shown in Figure 12), for categorization, there was a significant main effect of Block, $F(3, 66) = 7.72$, $p < .001$, $\eta_p^2 = .26$, and of Condition, $F(1, 22) = 26.71$, $p < .001$, $\eta_p^2 = .55$, but no significant interaction, $F = 1.12$. The systematic condition resulted in better categorization performance than the halfhalf condition. For individuation, for Block, $F(3, 66) = 24.97$, $p < .001$, $\eta_p^2 = .53$, for Condition, $F(1, 22) = 8.71$, $p < .01$, $\eta_p^2 = .28$, and the interaction, $F(3, 66) = 3.46$, $p < .05$, $\eta_p^2 = .14$. For individuation, the halfhalf condition was advantageous over the systematic condition, with this advantage increasing with longer training.

Finally, the planned comparison between the systematic condition with context from Experiment 2 and the halfhalf condition from Experiment 3 resulted in a similar pattern of results to the comparison with the systematic condition without context. For categorization, there was a significant main effect of Block, $F(3, 66) = 3.94$, $p < .05$, η_p^2

= .15, and of Condition, $F(1, 22) = 38.87$, $p < .001$, $\eta_p^2 = .64$, with an advantage for the systematic condition, and a significant interaction, $F(1, 22) = 2.98$, $p < .05$, $\eta_p^2 = .12$, due to the reducing difference in performance as training increased. For individuation, there was a significant main effect of Block, $F(3, 66) = 22.26$, $p < .001$, $\eta_p^2 = .50$, and of Condition, $F(1, 22) = 20.11$, $p < .001$, $\eta_p^2 = .48$, and a significant interaction, $F(3, 66) = 3.81$, $p < .05$, $\eta_p^2 = .15$. The advantage for the halfhalf condition over the systematic condition with context for individuation increased with additional training.

The results of both the halfhalf simulation and behavioural study supported our hypotheses. Including both arbitrary and systematic sound-form relationships in the words of an artificial language resulted in performance almost as effective as an entirely systematic mapping for categorization, and certainly better than an arbitrary mapping for this task. Critically, learning was more effective than an arbitrary mapping without additional context information in individuating meanings (and was substantially better than a systematic mapping for this task). Hence, it is beneficial for learning when the vocabulary consists of word forms that include systematic mappings related to category structure. This is because systematicity enables the similarities among categories of words to be discovered even if there is an absence of other contextual information providing cues to this category structure, and this category structure can then in turn be integrated with the remainder of the word which instantiates arbitrariness. Hence, some systematic component of the word may actually contribute to maximizing the information for individuation.

The question remains, however, whether words in natural language incorporate a similar division of labor between systematicity and arbitrariness. In our final study, we therefore conducted corpus analyses of English and French to determine whether the words in those languages have the same representational patterns as were used in the halfhalf model and behavioral experiment.

Corpus analyses: Arbitrary versus systematic mappings in natural language

Natural languages contain phonological cues to the grammatical category of a word, where grammatical category is one approach to considering general groupings of words in terms of meaning (Kelly, 1992; Monaghan et al., 2005). So, where in the word are we likely to find arbitrariness in terms of meaning and systematicity in terms of category? Speech processing is a fast, online, sequential process, consequently there is pressure on the beginning of a word to be distinct from other words, so that it can be uniquely identified as quickly as possible (Cutler et al., 1985; Lindell, Nicholls, & Castles, 2003). Hence, placing phonological information shared between many different words at the beginning of the word would impede the speed in which individuation could occur, due to online processing considerations. Placing this shared information towards the end of the word (reflecting the strong preference for suffixing over prefixing in the world's languages, Hawkins & Cutler, 1988) would provide systematicity for categorization purposes while also building into the word the kind of contextual cue that our previous studies have shown can facilitate meaning individuation when coupled with an arbitrary mapping.

Based on the results of Simulation 3 and Experiment 3, in terms of the categorization task we hypothesized that words in a natural language would contain more grammatical category information reflected in phonology at either the beginning or the end of the word. Due to processing and learning constraints (Cutler et al., 1985; St Clair et al., 2009), we hypothesized that this systematic information is most likely to occur at the end of the word. We tested categorization by determining the extent to which beginning or ending cues could predict the grammatical categories of words in the corpus using discriminant analysis. For individuation, we predicted that there would be greater differentiation in phonology in the region of the word that is least associated with grammatical category, i.e., at the beginning of the word. We tested this in several ways. First, we counted the number of beginnings versus endings. Note that this is a non-trivial division-of-labor, as an accurate categorization is more likely if there are multiple ways in which the word set can be distinguished. This can be tested explicitly by randomly reassigning the categories of words in the corpus and determining whether beginnings or endings more accurately categorize words. In this random case, we predict that if there are more beginnings then the beginnings will also be more accurate at classifying the random categories. Establishing that there is a dissociation between the number of distinct beginnings/endings and the part of the word most associated with categorization is evidence of the division of labor in the vocabulary supporting both categorization and individuation. As a second test of individuation, we measured the entropy (Shannon, 1951) of the distribution of words with each beginning and with each ending, predicting that entropy would be higher for beginnings than endings, reflecting greater variation in distribution of words according to their beginnings. Finally, we counted the number of words consistent with each beginning and ending. We predicted that there would be fewer words consistent with each beginning than with each ending.

We tested these hypotheses using corpus analyses, focusing on the distinction between nouns and verbs, while also seeking to determine whether the effect is cross-linguistic by analyzing both English and French. As previously discussed, one potentially powerful contributor to sound-category systematicity is morphology, which, in suffixing languages, tends to occur towards the end of words. Though the relationship between morphemes and grammatical categories is not straightforward (St Clair et al., 2009), and there are multiple phonological cues to category that are not related to morphology (e.g., Monaghan & Christiansen, 2008), we consider morphology to play an important role as a contextual cue in the vocabulary (Bybee, 1985), and that morphology is a mediator between surface forms of words and meaning (Seidenberg & Gonnerman, 2000).

Method

Corpus preparation. For English, we took the 5000 most frequent nouns and verbs, which were classified unambiguously in terms of grammatical category, from the English CELEX database (Baayen, Pipenbrock, & Gulikers, 1995). For French, we took the 5000 most frequent unambiguous nouns and verbs from the BRULEX database (Content, Mousty, & Radeau, 1990). Previous studies have focused on the noun/verb distinction in order to estimate the potential phonological information present in words for reflecting category membership (Kelly, 1992; Onnis & Christiansen, 2008), and we follow their lead here. There were 3,818 nouns and 1,182 verbs in the English corpus, and 3,657 nouns and 1,343 verbs in French.

Corpus analysis. To investigate the relationship between the phonology at the beginning of the word and grammatical category, we took as a cue the onset and nucleus of the first syllable (so for the word *penguin*, we used /pɛ/, and for the word ant, we used /æ/). For the end of the word cue, we took the nucleus and coda of the last syllable of the word (for penguin, we used /In/, and for ant, we used /ænt/). We chose the first and last few phonemes as participants have been found to be sensitive to the first phoneme for grammatical categorization (Arciuli & Cupples, 2007) and the first and last two letters of words have been shown to reflect stress patterns that in turn reflect grammatical category (Arciuli & Cupples, 2006). There were 566 distinct word beginnings, and 537 endings for English, and 455 beginnings and 167 endings for French.

The cues were entered into a discriminant analysis to determine how effectively the beginnings or endings of words related to the noun/verb distinction. As a baseline, we randomly reassigned the grammatical category labels to the words, and reran the analyses. These baseline analyses provide an empirical estimation of what chance-level performance would be if there were no division of labor for systematicity at the category level and arbitrariness at the individual word level[5].

Results and discussion

For English, the discriminant analysis on the beginning cues resulted in 71.6% correct classifications compared to 69.8% for the baseline. The ending cues correctly classified

---

[5] Note that the current analyses differ from those of St Clair et al. (2009) in that all word beginnings and endings were used in the analyses, rather than only morphemes, which enables a language-general approach to be taken in the corpus analyses. In addition, the analyses address the issue of categorization of the whole corpus, rather than determining for each morpheme whether it better predicts nouns or verbs.

92.0% of the words compared to 62.5% for the baseline. Using the beginning cue classifications as expected frequencies, and ending cue classifications as observed frequencies resulted in a significant difference, $\chi^2 = 1085.65$, $p < .001$. Hence, related to the categorization task, ending cues were significantly better for identifying the grammatical categories of words than beginning cues, indicating systematicity at the end of the word.

In terms of the individuation task, there were more distinct beginnings than endings in the corpus (566 compared to 537) and the random reassignment analysis for the baseline indicated that beginnings demonstrated greater distinctiveness in that they were able to predict an arbitrary random reassignment of categories better than the endings, as shown in the randomized baseline classification of beginnings as actual classifications versus endings as expected frequencies, $\chi^2 = 328.23$, $p < .001$.

Measures of entropy and the number of words consistent with beginnings versus endings resulted in a similar reflection of the individuation at the beginning rather than the end of the word. Entropy of word tokens per million for beginnings was 7.06 compared to 6.34 for endings, where high values indicate greater variability, and greater individuation. The mean number of words starting with each beginning was 8.83, lower than the mean number of words completed with each ending of 9.31, though this was not significantly different, $t(1101) = .25$, $p = .80$.

For French, the beginning cues resulted in 64.4% correct classification compared to 65.3% for the random baseline. For the ending cues, performance was again more accurate, with 89.6% correct classification compared to the 62.1% baseline, which was significantly higher than for beginnings, $\chi^2 = 1{,}382.02$, $p < .001$. Thus, as in English, ending cues are significantly better at categorizing words than beginnings. Related to the individuation task, again as for English there were more distinct beginnings than endings in the corpus, and a random categorization based on beginnings was significantly more accurate than categorization based on random endings, $\chi^2 = 57.52$, $p < .001$. Indeed, for French, actual beginnings were significantly less accurate than the random categorization, $\chi^2 = 321.14$, $p < .001$, though endings were significantly more accurate that the random classification, $\chi^2 = 2162.20$, $p < .001$. As for English, and indicating greater individuation at word beginnings than endings, entropy of frequencies per million was also greater over beginnings than endings, 7.08 and 4.66, respectively, and there were significantly fewer words on average with each beginning than with each ending, 10.99 and 29.94, $t(520) = 3.04$, $p < .005$.

The cues we have used in these analyses highlight the useful phonological information present in languages for determining grammatical category (Kelly, 1992). The use of the first and last few phonemes for each word reflects the findings of previous studies that have used the phonological form of the entire word (Monaghan et al., 2005), or just the first or last phoneme for categorization (Onnis & Christiansen, 2008). The corpus measures we employed contained morphological information because the analyses included multimorphemic words, but repeated analyses on monomorphemic nouns and verbs for phonological cues in the whole word resulted in similar effects (Monaghan et al., 2007), indicating that morphological markers to grammatical category are only a part of the contribution of phonological properties of words related to grammatical category.

Of particular interest in the current study, in both English and French the beginning of words provides more information about the identity of the word – providing more distinctiveness to assist in the identification of the unique word – yet the second half of the word is where greater systematicity can be observed between phonological forms and the general category for words. These results dovetail with those from the halfhalf condition in Simulation 3 and Experiment 3, where the words incorporated a similar division of labor between arbitrary and systematicity in the first and second half of the words, respectively.

The analyses have focused on only two languages, both of which are predominantly suffixing, but we predict a similar division of labor in any natural language – some part of the word will be systematic with respect to category, and another region of the word will bear an arbitrary relationship to category. Furthermore, we may make predictions about which type of affix is more likely to be prefixing or suffixing, in terms of how reliably it indicates the grammatical category of the word. Case markers and gender affixes are more likely to be suffixing than prefixing across the world's languages, for instance (Enrique-Arias, 2002; Hawkins & Gilligan, 1988), and these morphological markers are likely to be highly reliable indicators of the word's grammatical category. Additional factors may influence whether the systematic portion is at the beginning, at the end, or even in the centre of the word, but we have made a first step in showing how, in two languages, the dual role of the form-meaning mapping can be accommodated in natural language vocabularies.

General Discussion

The results of our computer simulations and human experiments combine to support our hypothesis of a division of labor between arbitrariness and systematicity inherent in the structure of words. The simulations and experiments we have presented here demonstrate the principles of learning using small sets of form-meaning mappings. We nonetheless suggest that the general principles demonstrated through laboratory-based studies such as these provide insight into the nature of the structure of languages, and raise hypotheses about the sorts of properties that we ought to observe in the form-meaning mappings of natural languages. Though the computational models were limited to small sets of words, the demonstration of similar learning effects in behavioural studies with relatively less-constrained human participants and in the large-scale corpus analyses of natural language support the scaling-up of these observations to claims about the structure of the whole vocabulary.

In the computational simulations of language learning, as well as the human studies of artificial language learning, the patterns of effects were consistent. For learning to individuate words, the optimal situation was contextual information plus arbitrary phonology-meaning mappings. Arbitrariness maximized the information present in each learning situation, and meant that identifying the correct referent for each word was accomplished most accurately. However, the arbitrary advantage for individuation was only observed when contextual information was also available. Indeed, an entirely arbitrary mapping with no contextual information even had a deleterious effect on learning the structure of the language, in terms of the category membership of words.

As word learning can occur with weak or absent contextual information, as in the case of low-frequency or first experiences with words, the advantage of arbitrary mappings for individuating meaning must be modulated by the potential usefulness of being able to generate, from the word’s phonology, a sense of its category. The halhalf model, where the mapping provided both arbitrariness for individual word meanings as well as systematicity at the category level in distinct regions of the word, illustrated one way in which these apparently conflicting requirements on the structure of the vocabulary can be met. The experimental results with these halfhalf stimuli demonstrated that this arrangement of information in the vocabulary was effective for learning both categorization and individuation. Indeed, across all the experimental conditions, the halfhalf condition resulted in the highest accuracy for individuation performance, and this was achieved without a substantial cost in terms of categorization performance. Our final corpus analyses of English and French further indicated that words in these two languages appear to be structured in a similar way: so as to incorporate the learning advantages of both arbitrary and systematic mappings at the beginnings and ends of words, respectively.

Our results raise the prediction that a similar division of labor will occur across most other suffix-dominant languages as well, because it is a consequence of learning constraints. Yet, languages with predominant prefixing or infixing are likely to show the distinction between the arbitrary and the systematic mappings at points other than the beginning and the end of the word. Prefixing languages, for instance, may well incorporate the systematic category-level information at the beginning of the word. We do not claim that the precise location of the arbitrary and systematic mappings are the same across all languages, but we do suggest that vocabularies in all languages will serve both to maximize information and generate a sense of the category from the phonological information within each word.

Morphological structure is one way in which this dual aim is met, and the halfhalf mapping in Simulation 3 and Experiment 3, in which both systematicity and arbitrariness was incorporated into the patterns to be learned, can be viewed as incorporating morphology at the end of the word. Common inflectional or derivational forms can provide a great deal of information about the grammatical category of the word, as shown in an analysis of the vocabulary in English child-directed speech by St Clair et al. (2009). Furthermore, they demonstrated that for English the majority of this systematicity occurred at the end of words – suffixes were more accurate cues to category than were prefixes. In the halfhalf model, this advantage for suffixes reflecting category was reflected in the vocabulary, with the form-category systematicity at the end of the word, and the arbitrary information for individuation at the beginning. However, the systematicity between form and category is not restricted only to morphology. Monaghan et al. (2005, 2007), building on earlier work by Kelly (1992), measured a range of phonological and prosodic cues that reflect grammatical category distinctions in child-directed speech, and found that these went far beyond morphological properties. In addition, the suffixing preference is not restricted to language materials, indicating that it reflects a more general learning bias (Hupp, Sloutsky, & Culicover, 2009).

A second predicted property of natural languages generated by the studies here is that systematic category-level information in the phonology of the word will be most beneficial for supporting word learning from arbitrary mappings when contextual information is not available, or is weak. In Simulation 2 and Experiment 2, we found that when contextual information was also available, the categorization task could be effectively solved when the mappings were arbitrary. Yet, when contextual information was not present then categorization accuracy from the arbitrary mappings was low. The advantage for arbitrary mappings over systematic mappings for individuation was found in our studies to be contingent upon contextual information also being present. Thus, for situations in natural language where contextual information is absent or unreliable, the arbitrary advantage will be sustained only if contextual information can also be embedded within the word's pronunciation. If words differ in terms of the strength or reliability of the contextual cues then one ought to observe differential inclusion of the systematic information within the word forms. In particular, words that are low-frequency have fewer opportunities for contextual information to be gathered to support either learning of their category or learning of their precise meaning. Also, words that occur in a wider range of contexts, again reducing the reliability of the contextual information to indicate their category, should also have more systematic information in terms of the phonology-category mapping for those words.

In analyses of child-directed speech, these patterns are precisely what is observed. For low-frequency words, the phonological cues that point to grammatical categories are much more reliable than for high-frequency words (Monaghan et al., 2005). Also, for words with more varied contexts, again we have observed that phonological cues to grammatical categories are more reliable. Verbs tend to occur in a wider range of contexts than nouns, and the systematicity in terms of phonology-category mappings is higher for verbs than for nouns across a range of languages (Monaghan et al., 2007). In Japanese, use of mimetics (or sound-symbolism) for verbs to assist in word-learning seems prominent. Mothers were more likely to use mimetic expressions for verbs when speaking to 18-20 month old children than when speaking to other adults (Nagamo, Imai, Kita, Haryu, & Kajikawa, 2006), and children aged 2 and 3 years old were able to use sound-symbolism to guide their learning of verbs (Imai, Kita, Nagumo, & Okada, 2008). We predict, based on the corpus analyses, that verb-learning is likely to benefit more than noun-learning from systematicity in the form-meaning mapping, though no direct comparison has yet been made. Indeed, Fitneva et al. (2009) provide indirect support for this prediction in the form of evidence showing that initial verb learning is facilitated more by phonological cues than noun learning.

The third simulation and experiment raise a further hypothesis for when arbitrary and systematic mappings will be respected by a particular word in natural language. When it is important to identify the word precisely, accurately, and swiftly, as in the potentially life-saving distinction between “Look! Kitten” and “Look! Leopard”, then arbitrariness is likely to be prominent in the mapping between the form and the meaning of the word. But when precise differentiations of meaning are not so critical then one may observe greater incidence, or greater refinement, in the systematicity of the mapping (e.g., as seen in expressives, cf. Gasser, Sethuraman, & Hockema, 2009). As it is not critically important to determine whether an object is large or huge, or whether it is tiny or teeny, under such conditions pressures for individuation to be supported by arbitrariness may be less critical than the requirement to convey the general sense of the word which can be most effectively supported by systematicity.

In summary, we have investigated the potential advantages of arbitrary and systematic mappings between word forms and their meaning, and have found that the optimal structure for language learning incorporates both arbitrariness and systematicity. From a learning perspective, the structure of the vocabulary in natural languages appears to strike a balance between supporting accurate individuation while also facilitating categorization. The fact that the vocabulary structure of natural languages, such as English and French, incorporate both of these learning properties suggests that the arbitrariness of the sign is not as complete as is conventionally thought. We have shown that effective learning of the vocabulary requires maximizing arbitrariness at the individual word level, but only when systematicity at the word-category level is signaled by contextual information—either in form of the distributional context of the word (Simulation/Experiment 2) or embedded within the word itself (Simulation/Experiment 3 and Corpus Analyses). These learning pressures may represent a universal feature of natural languages, that promote a division of labor between arbitrariness and systematicity in form-meaning mappings in order to serve both the individuation and the categorization functions important to language learning.

References

Arciuli, J. & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, *59*, 920-948.

Arciuli, J. & Cupples, L. (2007). Would you rather 'embert a cudsert' or 'cudsert an embert'? How spelling patterns at the beginning of English disyllables can cue grammatical category. In Schalley, A. & Khlentzos, D. (Eds.), *Mental states: Language and cognitive structure.* Amsterdam: John Benjamins.

Baayen, R.H., Pipenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bergen, B.K. (2004). The psychological reality of phonaesthemes. *Language, 80*, 290-311.

Black, A. W., Clark, R., Richmond, K., King, S., & Zen, H. (2004). *Festival Speech Synthesizer,* Version 1.95. CNRS, Edinburgh: University of Edinburgh.

Borges, J.L. (1942; 2000). John Wilkins' analytic language. In Weinberger, E., Allen, E., & Levine, S.J. (Eds.), *Borges: Selected non-fictions*, 229-232. New York: Penguin.

Brandone, A. C., Golinkoff, R. M., Pence, K. H., & Hirsh-Pasek, K. (2007). Action speaks louder than words: Young children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child Development, 78*, 1322 - 1342.

Brooks, P.B., Braine, M.D.S., Catalano, L., Brody, R.E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, *32*, 79-95.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Cassidy, K.W. & Kelly, M.H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, *30*, 348-369.

Cassidy, K.W. & Kelly, M.H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review*, *8*, 519-523.

Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences 31*, 489–558.

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds), *Language and thought: Interdisciplinary themes*, 162-183. Cambridge, UK: Cambridge University Press.

Content, A., Mousty, P., & Radeau, M. (1990). BRULEX: Une base de donnés lexicales informatisé pour le français écrit et parlé. *Anné Psychologique*, *90*, 551-566.

Corballis, M.C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The suffixing preference: A processing explanation. *Linguistics, 23*, 723–758.

Darwin, C. (1874). *The descent of man and selection in relation to sex* (2nd ed.). London: John Murray.

de Saussure, F. (1916). *Course in general linguistics*. New York: McGraw-Hill.

Dunn, L. & Dunn, L. (1997). *Peabody Picture Vocabulary Test III*. Circle Pines, MN: American Guidance Service.

Eco, U. (1995). *The search for the perfect language*. London: Blackwell.

Enrique-Arias, A. (2002). Accounting for the position of verbal agreement morphology with psycholinguistic and diachronic explanatory factors. *Studies in Language, 26,* 1–31.

Fitneva, S., Christiansen, M.H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*, 967-997.

Fudge, E. (1984). *English word-stress*. London: George Allen & Unwin.

Gasser, M. (2004). The origins of arbitrariness of language. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 434-439. Mahwah, NJ: Lawrence Erlbaum.

Gasser, M., Sethuraman, N., & Hockema. S. (2009). Iconicity in expressives: an empirical investigation. In J. Newman & S. Rice (Eds.), *Empirical and experimental methods in cognitive/functional research*. Stanford, CA: CSLI Publications.

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words, *Psychological Science, 20*, 347-353.

Greenberg, J. (1957). *Essays in linguistics*. Chicago: University of Chicago Press.

Hawkins, J.A. & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J.A. Hawkins (Ed.), *Explaining language universals*. Oxford: Blackwell.

Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua, 74*, 219-259.

Hinton, L., Nichols, J., & Ohala, J. (Eds.). (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Hockett, C. F. (1960). The origin of speech. *Scientific American,* 203, 89-96.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. (2000). Breaking the language barrier: An emergentist coalition model of word learning. *Monographs of the Society for Research in Child Development, 65*, Serial No. 262.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes, 24*, 876-909.

Hutchins, S. S. (1998). The psychological reality, variability, and compositionality of English phonesthemes. *Dissertation Abstracts International*, 59(08), 4500B. (University Microfilms No. AAT 9901857).

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54-65.

Kelly, M.H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*, 349-364.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114*, 19-28.

Lambon Ralph, M.A., & Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition, 13*, 928-948.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Langdon, M. (1984). Noise words in Guaraní. In Hinton, L., Nichols, J., & Ohala, J. (Eds.). (1994). *Sound symbolism,* 94-103. Cambridge: Cambridge University Press.

Lindell, A., Nicholls, M., & Castles, A.E. (2003). The effect of orthographic uniqueness and deviation points on lexical decisions: Evidence from unilateral and bilateral-redundant presentations. *Quarterly Journal of Experimental Psychology, 56A*, 287-307.

Mintz, T.H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition*, *30*, 678-686.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition, 96*, 143–182.

Monaghan, P. & Christiansen, M.H. (2006). Why form-meaning mappings are not entirely arbitrary in language. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives*, 139-164. Amsterdam: John Benjamins.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The Phonological Distributional Coherence Hypothesis: Cross-linguisitic evidence in language acquisition. *Cognitive Psychology, 55*, 259-305.

Monaghan, P. & Mattock, K. (2009). Cross-situational language learning: The effects of grammatical categories as constraints on referential labelling. *Proceedings of the Thirty First Annual Cognitive Science Society Conference*. Mahwah, NJ: Lawrence Erlbaum.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers, 36*, 402-407.

Newmeyer, F. J. (1998). *Language form and language function*. Cambridge, MA: MIT Press.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*, 181–186.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*, 1–16.

Onnis, L. & Christiansen, M.H. (2008). Lexical categories at the edge of the word. *Cognitive Science, 32,* 184-221.

Otis, K. & Sagi, E. (2008). Phonoaesthemes: A corpora-based analysis. In Love, B.C., McRae, K., & Sloutsky, V.M. (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. Boston, MA.

Seidenberg M.S. & Gonnerman, L.M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*, 353-361.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Shannon, C. E. (1951). Prediction and entropy of printed English. *The Bell System Technical Journal, 30*, 50-64.

Spector, F. & Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. *Developmental Psychology, 45*, 175-189.

St. Clair, M.C., Monaghan, P. & Christiansen, M.H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition, 116*, 341-360.

St. Clair, M.C., Monaghan, P. & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science, 33*, 1319-1329.

Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon, 3*, 259-278.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Boston, MA: Harvard University Press.

Ultan, R. (1978). Size-sound symbolism. In J. Greenberg (Ed.), *Universals of human language, volume 2 – Phonology*, 525-568. Stanford CA: Stanford University Press.

Wallis, J. (1699). *Grammar of the English language*, 5th edition. Oxford, UK: L. Lichfield.

Wilkins, J. (1668). *An essay towards a real character and a philosophical language*. London: Sa. Gellibrand.

Wittgenstein. L. (1953). *Philosophical investigations*. Oxford, UK: Blackwell.

Xu, F. & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science 10*, 288-297.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen month-old infants learn similar sounding words. *Developmental Science, 12*, 412-418.

Acknowledgements

Thanks to Nick Chater for helpful discussions over the theory in this paper, to three anonymous reviewers and to Amy Perfors for constructive and helpful comments on earlier drafts.

Table 1. Words used in Simulations 1 and 2 and Experiments 1 and 2.

Systematic Condition		Arbitrary Condition		Context words
Group 1 (fricative)	Group 2 (plosive)	Group 1	Group 2	
/fiʒ/	/ga:k/	/fiʒ/	/fɪʒ/	/wɛ/
/fɪʒ/	/gu:k/	/ʒɪf/	/ʒif/	/mə/
/ʒif/	/ka:g/	/fɪf/	/ʒiʒ/	
/ʒɪf/	/ku:g/	/ku:g/	/ga:k/	
/ʒiʒ/	/ka:k/	/ka:k/	/gu:k/	
/fɪf/	/gu:g/	/gu:g/	/ka:g/	

Table 2. Words used in Simulation 3 and Experiment 3.

Group 1	Group 2
/fiʒ/	/fik/
/ʒif/	/ʒig/
/ʒɪf/	/fɪk/
/ku:f/	/ga:k/
/ka:ʒ/	/gu:g/
/gu:ʒ/	/ka:g/

List of Figures

Figure 1. For Simulation 1, with no category marker words, accuracy of (A) same-category responses, and (B) individuation responses. Error bars indicate standard error of the mean in all figures.

Figure 2. Time taken to perfect performance for Simulation 1.

Figure 3. For Experiment 1, with no category-marker words, accuracy of (A) same-category responses, and (B) correct picture identifications.

Figure 4. For Simulation 2, with category-marker information, accuracy of (A) same-category responses and (B) individuation responses.

Figure 5. Time taken to perfect performance for Simulation 2.

Figure 6. For Experiment 2, with category-marker words, accuracy of (A) same-category responses, and (B) correct picture identifications.

Figure 7. First two Principal Components of hidden unit representations for arbitrary and systematic models with and without context information. Each point indicates one of the words used in the artificial language learning experiments, black squares indicate words from one category, and white circles indicate words from the other category. The left and

right columns show the representations after 10 and 40 epochs of training, respectively AN: arbitrary no context; AC: arbitrary with context; SN: systematic no context; SC: systematic with context.

Figure 8. Mean Euclidean distance among hidden unit representations of the same category (within) and of different categories (between). AN: arbitrary no context; AC: arbitrary with context; SN: systematic no context; SC: systematic with context.

Figure 9. Minimum Euclidean distance among hidden units representations for (A) arbitrary, and (S) systematic models.

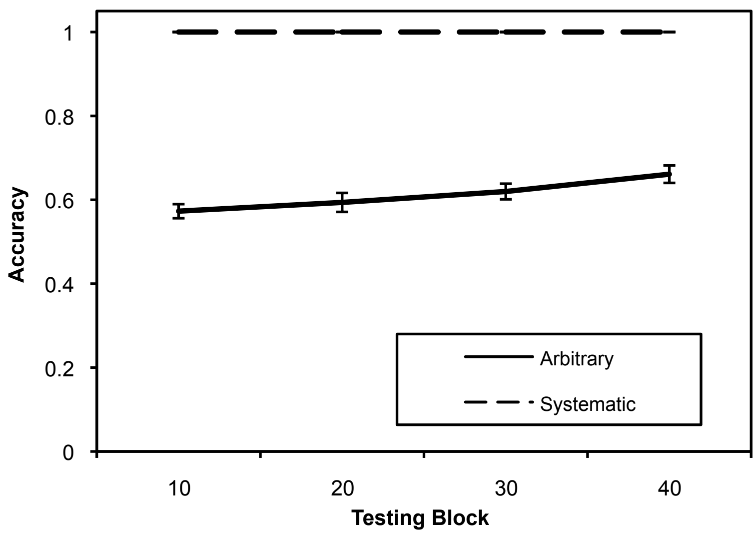
Figure 10. Accuracy of (A) categorization and (B) individuation responses for halfhalf condition in Simulation 3 without context, compared to results from Simulation 1.

Figure 11. Accuracy of (A) categorization and (B) individuation responses for halfhalf condition in Simulation 3 with context, compared to results from Simulation 2.

Figure 12. Accuracy of (A) categorization and (B) individuation responses for halfhalf and arbitrary conditions in Experiment 3.

Figure 1.

A


1
0.8
0.6
0.4
0.2
0
Accuracy
Arbitrary
Systematic
10
20
30
40
Testing Block

B

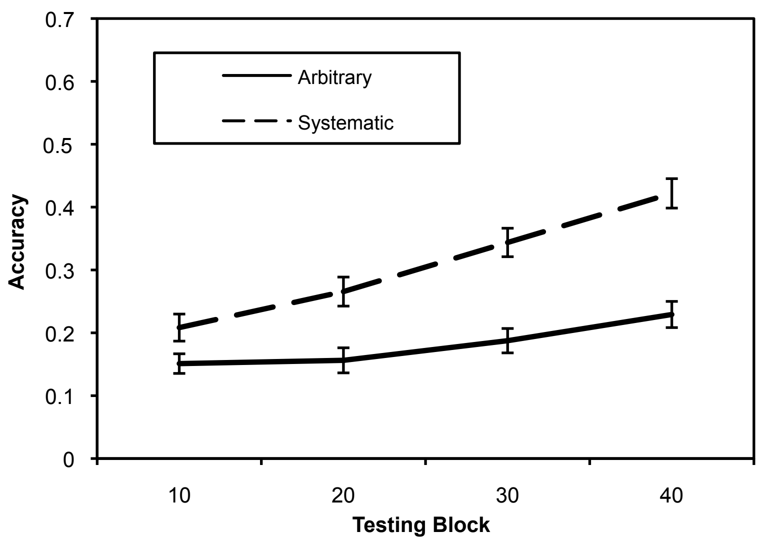

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
Arbitrary
Systematic
10
20
30
40
Testing Block

Figure 2.

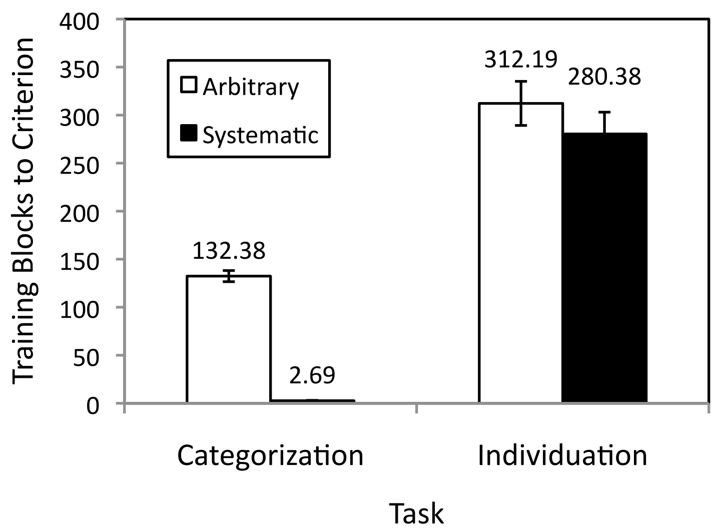
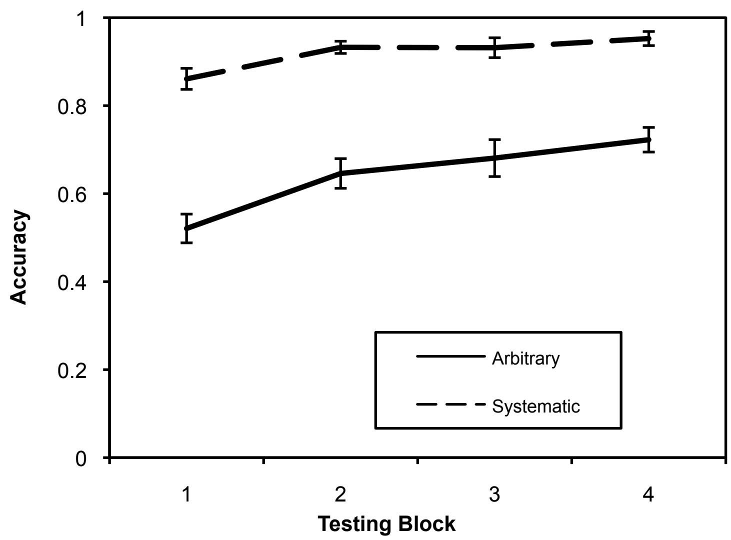

400
350
300
250
200
150
100
50
0
Training Blocks to Criterion
Arbitrary
Systematic
132.38
2.69
312.19
280.38
Categorization
Individuation
Task

Figure 3.

A


1
0.8
0.6
0.4
0.2
0
Accuracy
1
2
3
4
Testing Block
Arbitrary
Systematic

B

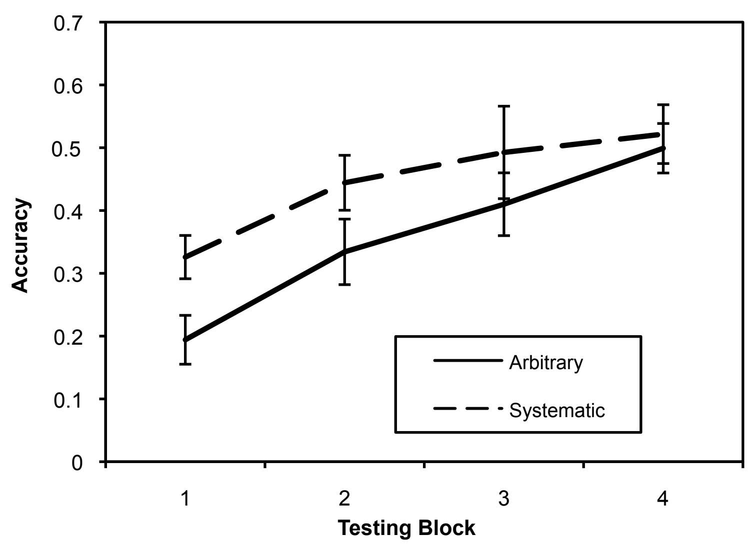
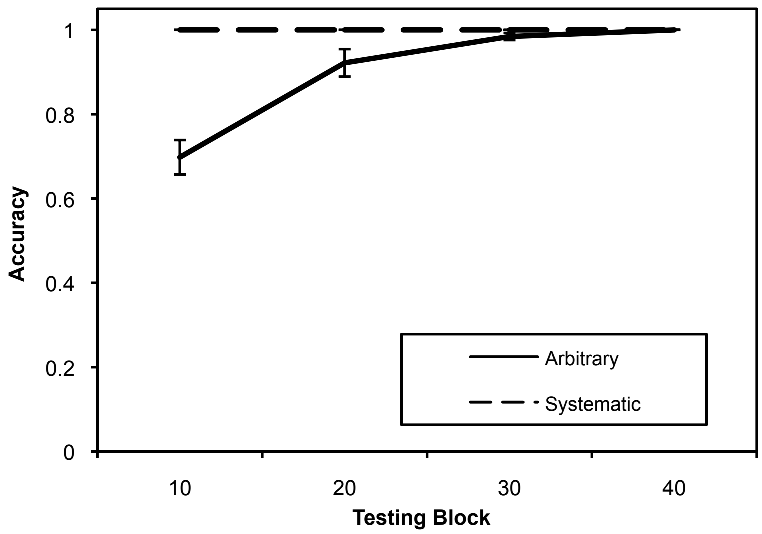

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
1
2
3
4
Testing Block
Arbitrary
Systematic

Figure 4.

A


1
0.8
0.6
0.4
0.2
0
Accuracy
Arbitrary
Systematic
10
20
30
40
Testing Block

B

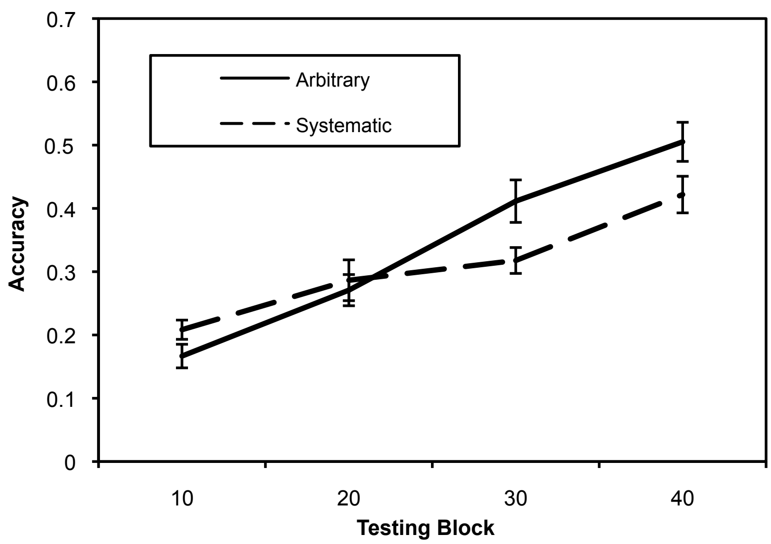

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
Arbitrary
Systematic
10
20
30
40
Testing Block

Figure 5.

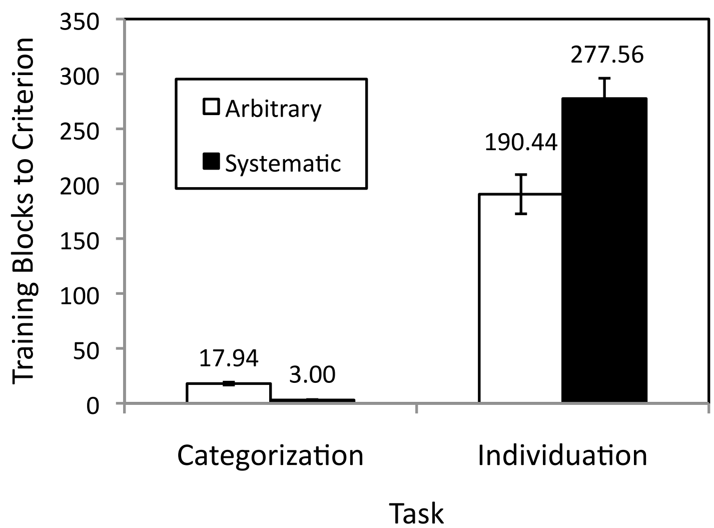
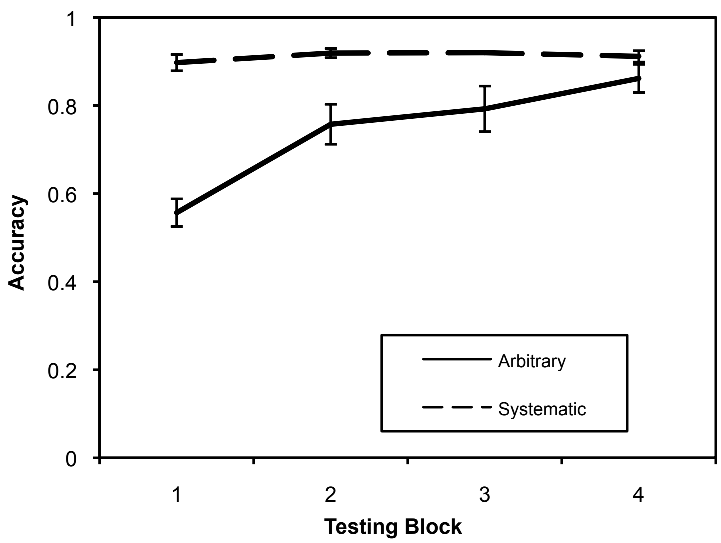

350
300
250
200
150
100
50
0
Training Blocks to Criterion
Arbitrary
Systematic
17.94
3.00
190.44
277.56
Categorization
Individuation
Task

Figure 6.

A


1
0.8
0.6
0.4
0.2
0
Accuracy
1
2
3
4
Testing Block
Arbitrary
Systematic

B

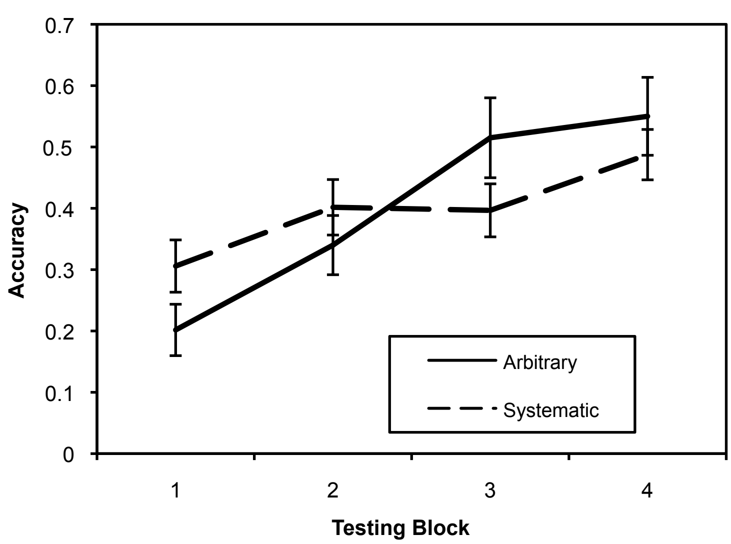

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
1
2
3
4
Testing Block
Arbitrary
Systematic

Figure 7.

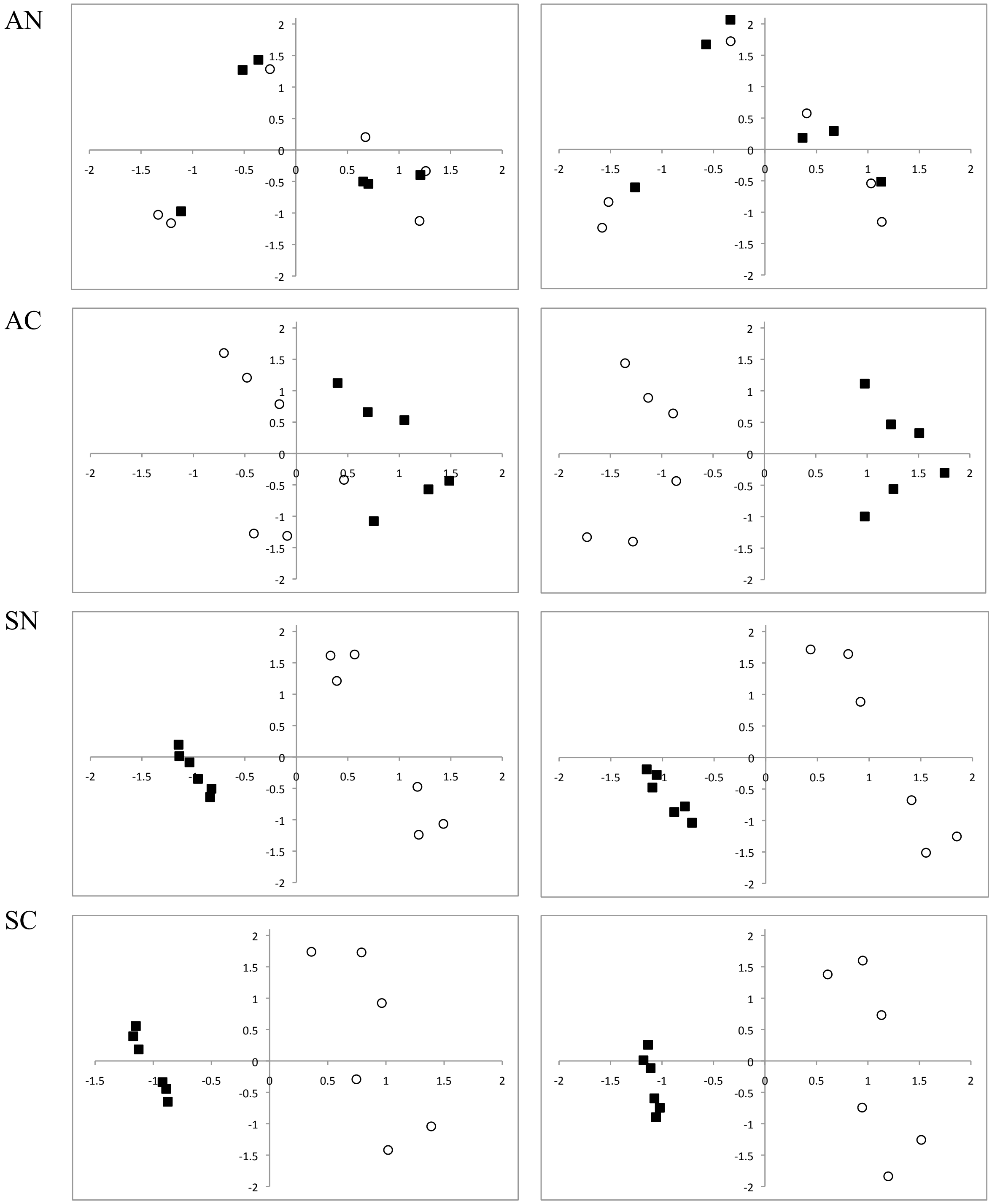


Figure 8.

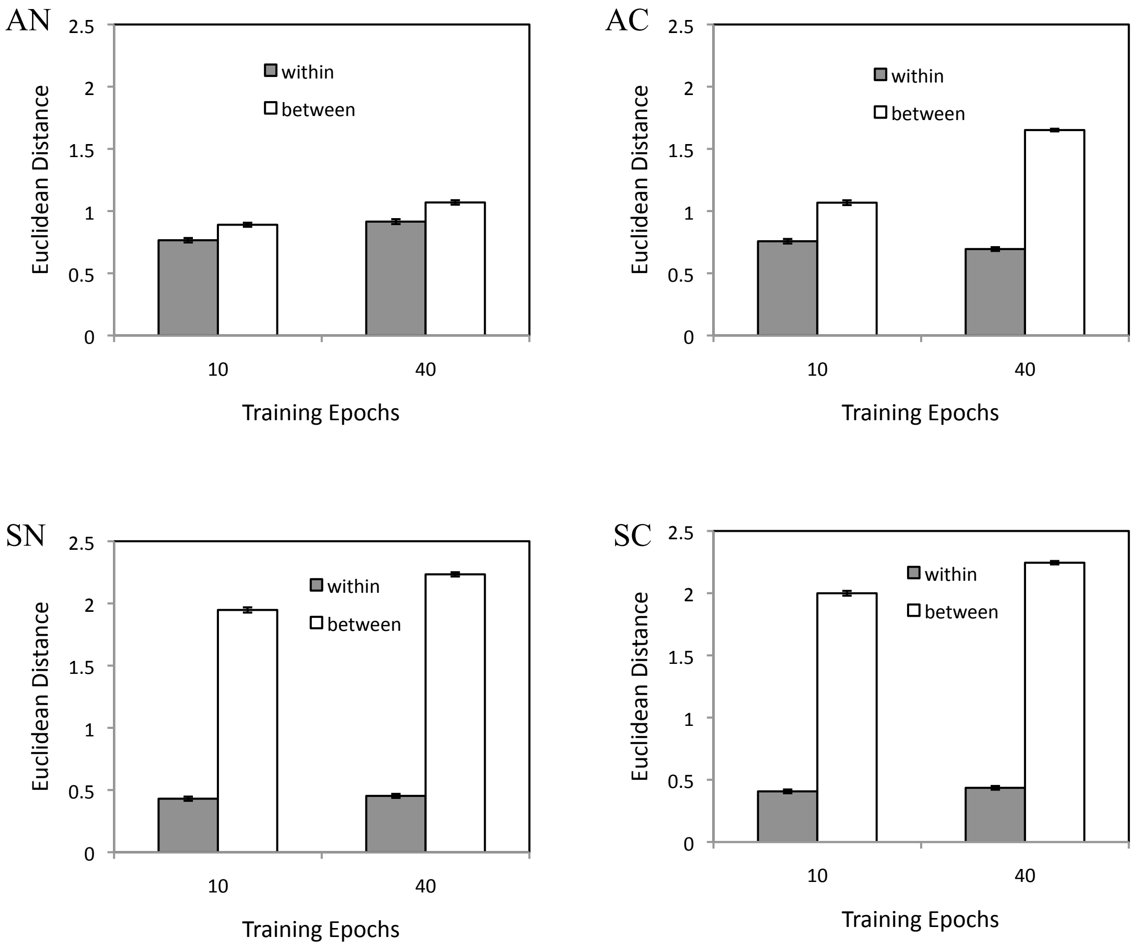
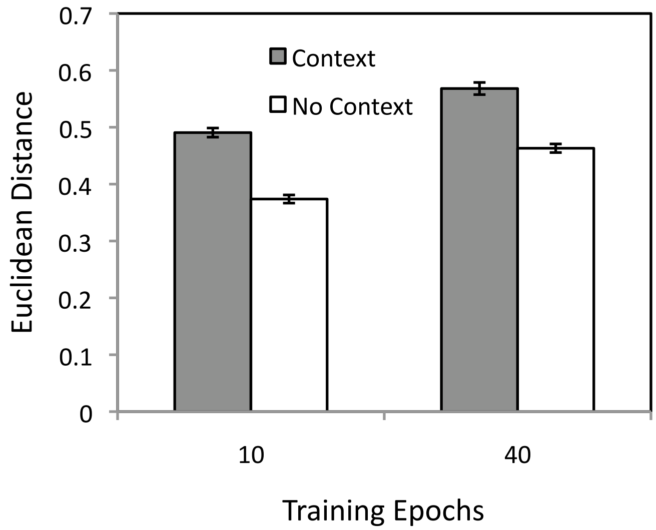

AN
2.5
2
1.5
1
0.5
0
Euclidean Distance
within
between
10
40
Training Epochs
AC
2.5
2
1.5
1
0.5
0
Euclidean Distance
within
between
10
40
Training Epochs
SN
2.5
2
1.5
1
0.5
0
Euclidean Distance
within
between
10
40
Training Epochs
SC
2.5
2
1.5
1
0.5
0
Euclidean Distance
within
between
10
40
Training Epochs

Figure 9

A


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Euclidean Distance
Context
No Context
10
40
Training Epochs

S

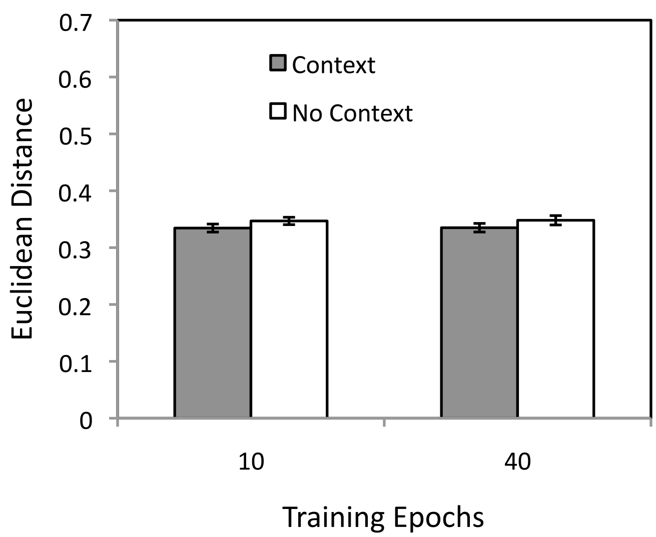
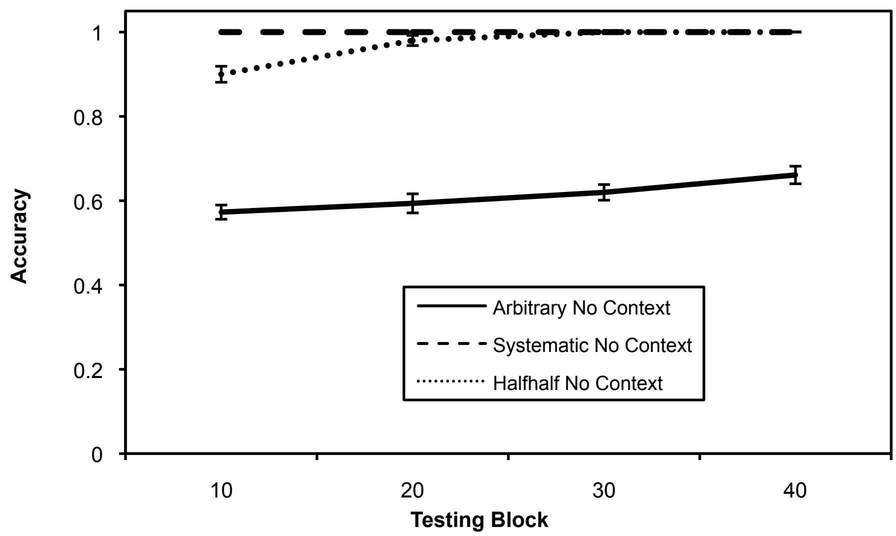

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Euclidean Distance
Context
No Context
10
40
Training Epochs

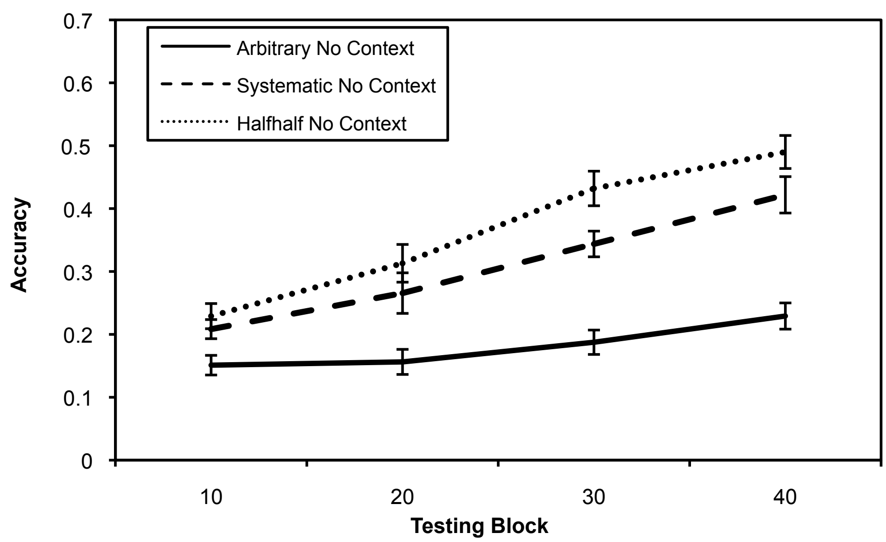
Figure 10

A



1
0.8
0.6
0.4
0.2
0
Accuracy
10
20
30
40
Testing Block
Arbitrary No Context
Systematic No Context
Halfhalf No Context

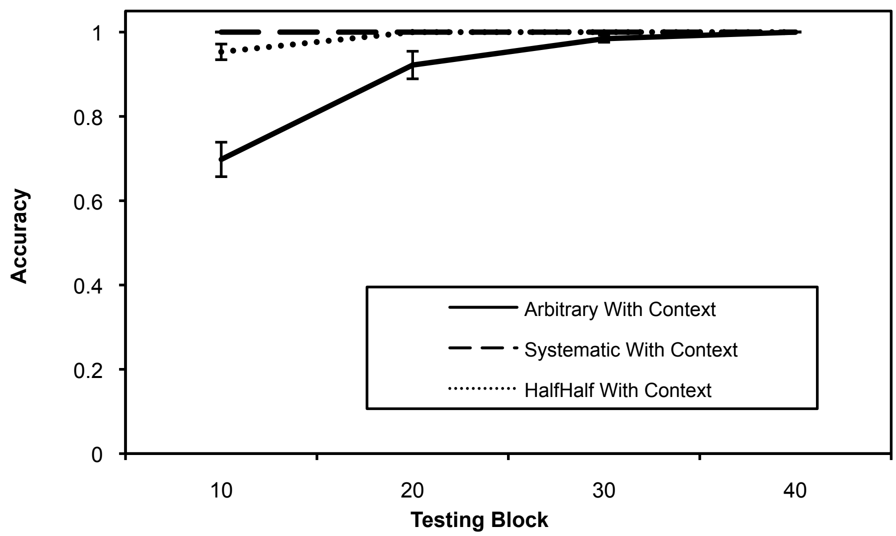
B



0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
10
20
30
40
Testing Block
Arbitrary No Context
Systematic No Context
Halfhalf No Context

Figure 11

A


1
0.8
0.6
0.4
0.2
0
Accuracy
10
20
30
40
Testing Block
Arbitrary With Context
Systematic With Context
HalfHalf With Context

B

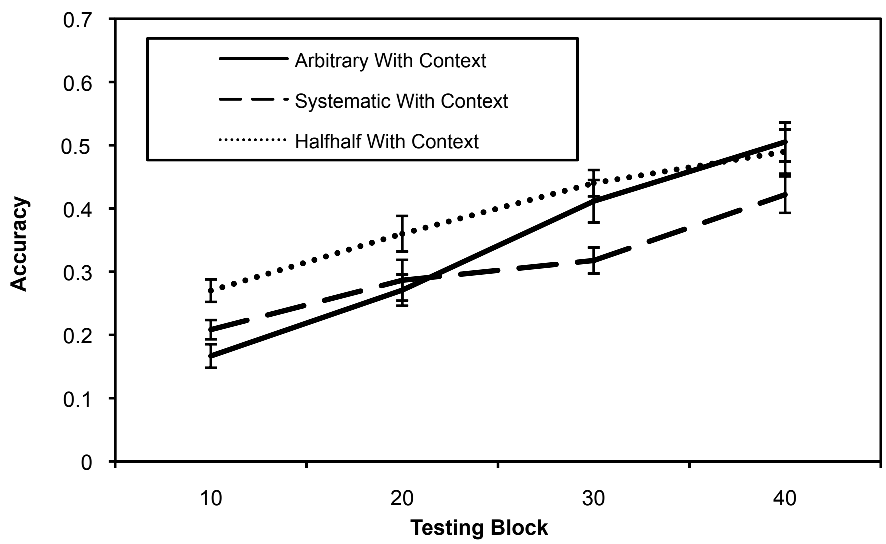
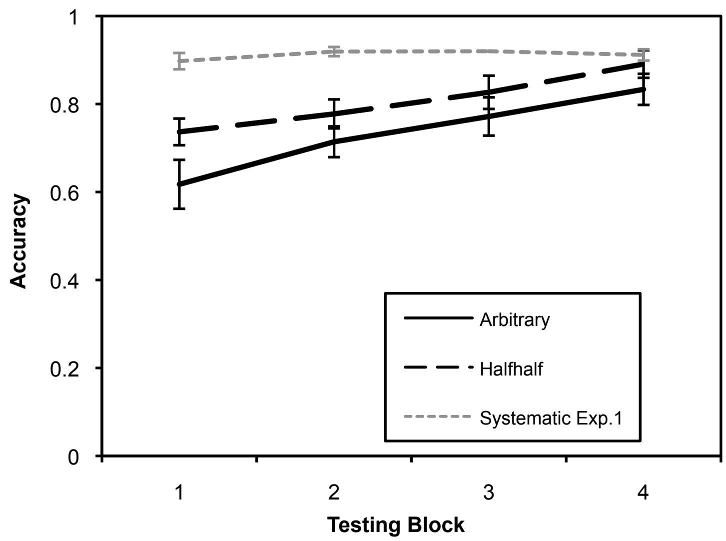
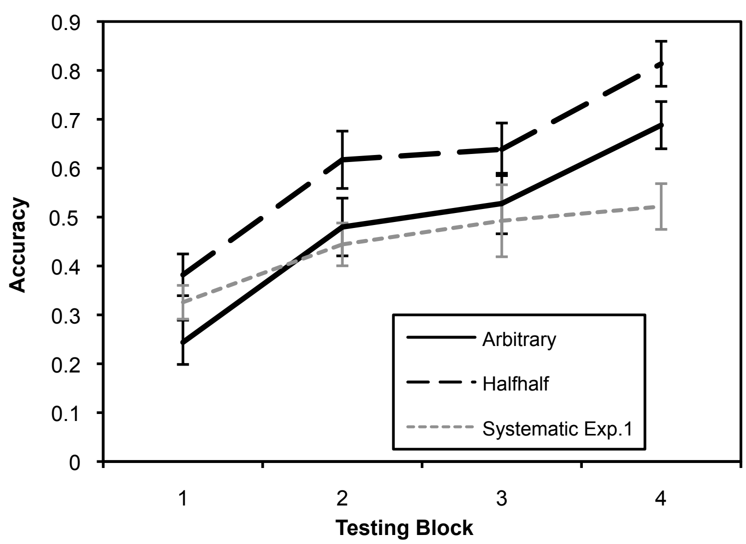

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
10
20
30
40
Testing Block
Arbitrary With Context
Systematic With Context
Halfhalf With Context

Figure 12

A


Accuracy
1
0.8
0.6
0.4
0.2
0
1
2
3
4
Testing Block
Arbitrary
Halfhalf
Systematic Exp.1

B


Accuracy
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
Testing Block
Arbitrary
Halfhalf
Systematic Exp.1

Appendix

List of pictures from which 6 actions and 6 objects were selected randomly for each participant.

Action pictures: somersaulting, trampolining, clapping, painting, running, swimming, climbing, crying, feeding, dancing, swinging, mopping, tickling, pulling, chopping, tearing

Object pictures: spoon, key, truck, wagon, lamp, goat, drum, feather, window, arrow, nest, broom, pineapple, fish-hook, pail, frog