PAPER

Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns

LouAnn Gerken,[1] Frances K. Balcomb[2] and Juliet L. Minton[1]

1. Department of Psychology, University of Arizona, USA
2. Neuropsychiatry Program, University of Pennsylvania Medical Center, USA

Abstract

Every environment contains infinite potential features and correlations among features, or patterns. Detecting valid and learnable patterns in one environment is beneficial for learners because doing so lends predictability to new environments where the same or analogous patterns recur. However, some apparent correlations among features reflect spurious patterns, and attempting to learn the latter costs time and resources with no advantage to the learner. Thus, an efficient learner in a complex environment needs to devote more attention to input that reflects a real and learnable pattern than to input that reflects a spurious or ultimately unlearnable pattern. However, in order to achieve such efficiency in the absence of external feedback, learners need to have an implicit metric of their own learning progress. Do human infants have such a metric? Data from two experiments demonstrate that 17-month-olds attend longer to learnable vs. unlearnable linguistic grammars, taking more trials to habituate and more overall time to habituate for grammars in which a valid generalization over input stimuli can be made. These data provide the first evidence that infants have an implicit metric of their own learning progress and preferentially direct their attention to learnable aspects of their environment.

Introduction

Learning about the world entails finding patterns among experiences. Pattern detection allows organisms to generalize over their experiences and thereby predict new experiences that might arise in the future. For example, having learned one structure allows human adults and infants to more readily detect a similar structure in novel input (Lany, Gómez & Gerken, 2007). Human infants are very adept at generalizing over a variety of visual and auditory patterns without receiving external feedback (Gerken, Wilson & Lewis, 2005; Gómez & Gerken, 1999; Marcus, Vijayan, Rao & Vishton, 1999; Needham, Dueker & Lockhead, 2005; Quinn & Bhatt, 2005). However, in the real world, potential patterns differ in their ease of detection and validity. For example, if you are given the numbers 42, 56, 14, and 21, you might predict that if there are more numbers in this set, another multiple of 7 such as 70 might be among them. In contrast, if you are given 42, 56, 15, and 23, you might find it more difficult to predict additional numbers, because the four examples were selected so as not to reflect obvious arithmetic patterns. Of course, other types of patterns might be observed in the numbers, but it is likely to take some time and effort to induce them, if they do indeed exist, perhaps reducing the value of doing so.

Are learners able to determine whether or not they are making progress on inducing a pattern and thereby devote learning time and resources efficiently? Or do they devote a set amount of time and resources to each potential pattern, sometimes labouring in vain over unlearnable input (Nelson & Leonesio, 1988)? Data suggest that adult humans have some ability to determine how they are progressing on a learning problem and to dedicate study resources appropriately (Metcalfe, 2009; Metcalfe & Kornell, 2005). These data support a learning model in which adults attempt to achieve a maximal rate of learning. Low or zero learning rates occur when a problem is either so difficult that no learning can occur or the problem has been fully mastered. Thus, information can be discarded as a focus of learning either because it is already known or because it is not knowable despite additional efforts. Explicit metacognitive abilities of this sort are typically not attributed to children before the early school years (Cultice, Somerville & Wellman, 1983; Flavell, Green & Flavell, 2000; Lockl & Schneider, 2002). However, recent data suggest that 3-year-olds have tacit access to their own knowledge states (Balcomb & Gerken, 2008). In developmental research that specifically addresses the deployment of attention to different types of learning situations, 4-year-olds show greater exploration of toys for which they have ambiguous vs. unambiguous evidence about causality (Schulz &

Address for correspondence: LouAnn Gerken, Department of Psychology, 1503 E. University Blvd., University of Arizona, Tucson AZ 85721-0068, USA; e-mail: gerken@u.arizona.edu

Bonawitz, 2007). In addition, many studies of infants assume an ability to determine whether visual or auditory information presented to them in the laboratory is already known to them or something new to be learned (Hunter & Ames, 1988). In these studies, infants are allowed to experience a stimulus for as long as they show interest. When interest flags (when the infant has habituated), it is assumed that the infant has learned as much as subjectively possible. For example, infants who were interrupted while exploring a set of toys were more likely to continue exploring the original set of toys when given the choice of this set or a new set, than infants who were allowed to play with the toys until they had begun to turn their attention elsewhere. The latter group was more likely to choose a new set of toys when given a choice (Hunter, Ames & Koopman, 1983).

The studies just outlined suggest that infants and young children differentially attend to already learned (or unambiguous) vs. unlearned (or ambiguous) domains. Such differential attention might be due merely to distinguishing familiar from unfamiliar patterns. Alternatively, the existing results might reflect a more sophisticated internal metric of learning like the one proposed for adults (Metcalfe, 2009; Metcalfe & Kornell, 2005). This metric allows learners to implicitly track whether or not they are making progress on a problem, even while they are in the process of trying to master it. Such a metric of learning would allow learners to distinguish learnable vs. unlearnable domains (see following paragraph for definitions of these terms). Although an internal metric of learning progress has been heretofore unexplored in infants, there are a priori reasons to predict its existence. While it might be adaptive to persist in attempts to discern a pattern that is learnable but not yet learned, it is almost certainly maladaptive to persist at looking for a pattern in a random set of input. Therefore, it is difficult to imagine that young learners could be as successful as they are if they devote the same amount of time and resources to learnable and unlearnable patterns. Furthermore, if implicit awareness of learning progress serves to direct attention toward a particular learning domain, we could explain why infants are so readily able to learn a variety of non-functional patterns in the laboratory when no external reinforcement is provided. That is, a feeling of learning is in itself reinforcing. The two experiments reported here test the hypothesis that infants have an implicit metric of their learning progress, which drives attention toward learnable and away from unlearnable linguistic patterns (i.e. grammars).

Before we turn to the experiments, however, we must elaborate further on the notion of a learnable problem that we are using here by making two distinctions. First, we want to distinguish between learning that entails merely the encoding and recall of stimuli with the learning that is the focus of our research: A learnable problem is one that allows the learner to find deeper level patterns or generalizations in the stimuli themselves. For example, a set of Russian words that can be divided into two gender classes, each class taking one of two case endings, is an example of a learnable problem under this definition. One way of operationalizing the generalization notion of learning comes from the concept of data compression in information theory: a compressed representation of data is one that uses fewer bits than required to represent the data themselves (e.g. Shannon & Weaver, 1949). Thus, a learnable problem of the sort that we are studying here is one that allows data compression. Although we will not engage in formal testing of our input stimuli to determine if they are indeed compressible, it is important to acknowledge that a formal definition exists. The idea that finding a generalization (a data compression algorithm) in a set of data is somehow interesting or affectively positive can be found in several sources, both popular and academic, including Johnson (1995) and Gopnik (2000). This idea is also consistent with developmental data discussed above showing a link between attention/interest and extracting a generalization over a set of observations about the cause of those observations (e.g. Schulz & Bonawitz, 2007).

The second distinction concerning learning that we need to make is that a problem can be defined as learnable both objectively and subjectively. Objectively, a learnable problem is one that allows for generalization or data compression. However, it may be the case that an objectively learnable problem cannot be learned by a particular mind and is therefore not subjectively learnable. In the work presented here, we define a learnable problem as one that is both objectively and subjectively learnable. In contrast, an unlearnable problem can be either objectively unlearnable (the data are not compressible) or subjectively unlearnable (the data are compressible, but the learner is unable to determine how).

One domain for which we have evidence of both objective and subjective learnability is the inference of linguistic categories from a distribution of words in the input. Consider the six feminine and six masculine Russian words shown in Table 1. The feminine words take the endings *oj* and *u*, and the masculine words take the endings *ya* and *yem*. In addition, some of the words in each set have a unique category marker: The stems of the first three feminine words end in *k*, and the comparable stems in the first three masculine words end in *tel*. Studies of adults and 17-month-olds who listened to the words in Table 1 with a subset of items withheld show that they can discriminate the withheld test items (e.g. *vannoj*) from equally unfamiliar ungrammatical items

Table 1 *Stimuli for the learnable grammar used in Experiment 1. Bolded words were not presented in the experiment*

Feminine words					
polkoj	rubashkoj	ruchkoj	**vannoj**	knigoj	korovoj
polku	rubashku	ruchku	vannu	knigu	**korovu**
Masculine words					
uchitelya	stroitelya	zhitelya	**medvedya**	kornya	pisarya
uchitelyem	stroitelyem	zhitelyem	medvedyem	kornyem	**pisaryem**

(e.g. *vannya*). Thus, the category system in Table 1 is learnable.

Now consider the very similar Russian feminine and masculine words in Table 2. Here the feminine and masculine words appear with the same endings as the words in Table 1 (*oj* and *u* for feminine words and *ya* and *yem* for masculine words), but the unique category markers *k* and *tel* are not present. Although this learning problem is objectively learnable, in the sense that it can be compressed into a statement about categories and case endings, several decades of research with adults and infants have shown it to be subjectively unlearnable (Braine, 1963a; Gerken, Wilson, Gomez & Nurmsoo, 2009; Gerken *et al.*, 2005; Smith, 1966). The only information that learners appear to extract from this unlearnable problem is that each word stem is followed by an ending, but knowing one ending does not allow the prediction of another (Smith, 1966). Thus, it appears that the additional word stem marking of the categories must be present on at least some category members for the problem to be learnable by infants and adults (Braine, 1963b; Gerken *et al.*, 2009; Gerken *et al.*, 2005; Mintz, 2002).

We know from previous research that 17-month-olds showed successful learning for stimuli like those in Table 1, but not for stimuli like those in Table 2, when the input was presented for an equal and fixed amount of time (2 min.), independent of infants' interest (Gerken *et al.*, 2005). The question addressed in the current research is whether infants of the same age who are given the choice of how long to attend to input would differentiate a problem that they can solve from one they cannot, attending longer to input from the learnable pattern than to one that is either subjectively unlearnable (Experiment 1) or objectively unlearnable (Experiment 2). If they do, we would have some evidence that very young learners can distribute attention according to implicitly perceived progress on generalizable pattern detection.

Experiment 1

Experiment 1 asked whether 17-month-olds would take longer to habituate to the learnable stimuli in Table 1 than the subjectively unlearnable stimuli in Table 2.

Methods

Participants were 30 infants ranging in age from 16 months, 2 weeks to 17 months, 2 weeks with no family

Table 2 *Stimuli for the unlearnable grammar. used in Experiment 1. Bolded words were not presented in the experiment*

Feminine words					
lapoj	malinoj	ruchkoj	**vannoj**	knigoj	korovoj
lapu	malinu	ruchku	vannu	knigu	**korovu**
Masculine words					
tramvaya	iulya	zhitelya	**medvedya**	kornya	pisarya
tramvayem	iulyem	zhitelyem	medvedyem	kornyem	**pisaryem**

history of speech problems or language delays (aside from minor pronunciation issues), no current or recent ear infections, primarily exposed to English (with no more than 10 hours a week of any other language exposure), born at full term (at least 37 weeks gestation), and a weight of at least 5 lb 8 oz at birth. Fifteen infants participated in the learnable condition (12 females) and 15 in the unlearnable condition (10 females). Six additional infants were tested but not included because of equipment failure (one), or they cried or became fussy before meeting the criterion for habituation (two in the learnable and three in the unlearnable condition).

Materials were the 20 non-bolded words of Russian from Table 1 (learnable condition) and Table 2 (subjectively unlearnable condition), spoken by a fluent speaker of Russian. Four different random orders of the stimuli from each condition were created.

Methods: During the experiment, infants were seated in a highchair facing a wall with a white screen between two speakers. The accompanying adult was asked to sit quietly in an adjacent chair, and not to speak, gesture, or otherwise direct the infant's attention. A projector mounted above the seating displayed images on the screen, and a video camera mounted above the screen allowed the experimenter to track and record the infant's eye gaze. The program Habit X 1.0 (Cohen, Atkinson & Chaput, 2004) was used to present infants with the four auditory stimuli of the relevant condition, with the presentation randomized in blocks of four trials. The auditory stimuli were paired with a static image of a red and black checkerboard during the habituation trials. Presented in between each of the habituation trials was an animated baby face sucking on a pacifier, which served to draw the infant's attention back to the screen for the presentation of the next trial. Habituation trials began when the infant looked at the screen and ended when the infant looked away for 2 or more seconds. Habituation was considered to have occurred when total looking time over a fixed window of four trials was less than 50% of looking time across the first four trials.

Results

Our prediction was that infants in the learnable condition would listen longer than infants in the unlearnable condition. Two dependent measures of listening time were used: the number of trials to habituate (Figure 1a) and the overall time to habituate (Figure 1b). The number of habituation trials exhibited by infants presented with a grammar previously established to be learnable ranged from 8 to 48 trials with a mean of 22.93 (*SD* = 11.6), while the number in the unlearnable condition ranged from 8 to 40 habituation trials with a mean of 14.93 (*SD* = 8.88). A *t*-test showed this difference to be significant (*t*(28) = 2.13, p < .05, two-tailed). With respect to overall time to habituate (Figure 1b), infants exposed to the learnable grammar
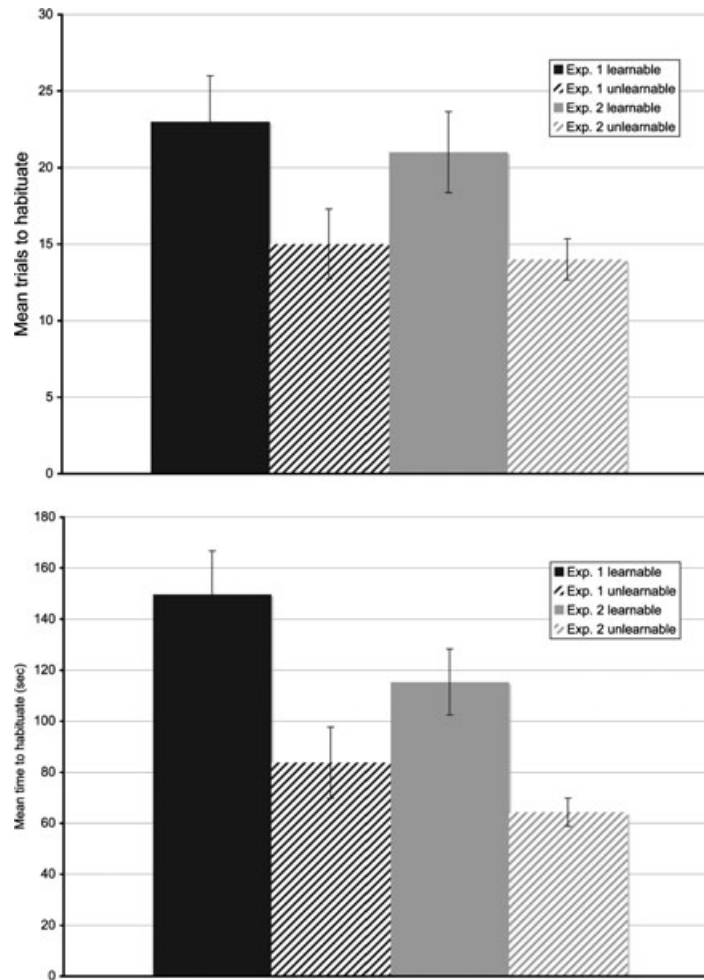
Figure 1   *1a (top) and 1b (bottom). Trials to habituate and time to habituate and standard errors in first and second experiments.*

showed a range of times between 90 and 244 sec. with a mean of 149.71 (SD = 63.07), while infants in the un-learnable condition showed a range of times between 25 and 232 sec. with a mean of 83.7 (SD = 52). A *t*-test also showed the difference in total time to habituate to be significant ($t(28)$ = 3.13, $p < .01$, two-tailed). Thus our main prediction was borne out for both dependent measures.

We take the data from Experiment 1 to indicate that infants show greater interest in the grammar that other infants and adults tested with a different method were able to learn than one that other infants and adults were unable to learn. An alternative and less theoretically interesting interpretation of the data is that some low level feature made the stimuli from the learnable condition intrinsically more interesting. For example, these stimuli contained more repeated elements (e.g. the *k* and *tel* word endings). Experiment 2 was designed to rule out such an interpretation, as well as to conceptually replicate the basic findings of Experiment 1 by comparing infants' interest in a grammar that should be both objectively and subjectively learnable vs. an objectively unlearnable grammar.

Experiment 2

Experiment 2 asked whether 17-month-olds would take longer to habituate to a new set of putatively learnable stimuli than stimuli for which no compression is possible, thereby making the stimuli objectively unlearnable. The new learnable stimuli were created from the stimuli in Table 1 by placing the masculine endings on the feminine words and vice versa. We call this new grammar 'anti-Russian'. The objectively unlearnable stimuli were created from the union of the stimuli in Tables 1 and anti-Russian.

Although we did not confirm its learnability, we reasoned that anti-Russian, because it has the same basic structure as the learnable stimuli in Experiment 1 (including unique category marking), should be both objectively and subjectively learnable and therefore support infants' attention, just as the learnable version of Russian did in the first experiment. In contrast, the union of Russian and anti-Russian is objectively unlearnable, and infants presented with this grammar should show significantly shorter listening times than

infants exposed to anti-Russian. Importantly, if this pattern of results is found, it cannot be due to low level stimulus properties of either Russian or anti-Russian. Rather, if infants prefer both Russian and anti-Russian to the union of the two, this preference is most likely due to the objective unlearnability of the latter.

Methods

Participants were 30 infants ranging in age from 16 months, 2 weeks to 17 months, 3 weeks, subject to the same inclusion criteria as those in Experiment 1. Fifteen infants participated in the learnable condition (10 female) and 15 in the unlearnable condition (seven female). Twenty-three additional infants were tested but not included because they cried or became fussy before meeting the criterion for habituation (10 in the learnable and 13 in the unlearnable condition).[1]

Materials for the learnable grammar had the same type of structure as the one shown in the learnable grammar of the first experiment (Table 1), including the unique category markers *k* and *tel*; however, the two case endings that had previously occurred on feminine words (*oj* and *u*) were placed on the masculine words and vice versa. The *k* and *tel* endings were left on the original Russian stems. As noted above, we refer to this new grammar as 'anti-Russian'. The objectively unlearnable grammar was the union of learnable Russian shown in Table 1 and the anti-Russian version of that grammar. All stimuli for Experiment 2 were naturally produced, not cross-spliced from Experiment 1. Although both Russian and anti-Russian should be learnable on their own, their union creates a system in which the previously learnable internal structure is lost, while maintaining equivalence of feature complexity across the two grammars. For example, while only some word stems end in *k* and others in *tel*, all words can occur with the endings *oj*, *u*, *ya* and *yem* in the union of Russian and anti-Russian. Note that because the unlearnable stimuli in Experiment 2 are the union of Russian and anti-Russian, there are twice as many words as in any of the other conditions. Because the goal of Experiment 2 was to rule out any low level property of the input that might cause infants to attend more to the learnable stimuli in Experiments 1 and 2, we didn't want to take the chance of leaving out particular stimuli that might be driving the effect. Therefore, we felt that this potential confound was warranted.

[1] The number of infants who failed to habituate in Experiment 2 is almost four times that for Experiment 1 (total time to habituate was also longer in Experiment 1, but not trials to habituate). We attribute this difference to the fact that it is nearly impossible for a human talker to produce ungrammatical items as fluently as grammatical items. Thus, infants may have detected some subtle lack of fluency in the anti-Russian items, which dampened their interest in those stimuli (Soderstrom & Morgan, 2007). Note that this possible fluency difference cannot account for their longer listening times for anti-Russian than the union of Russian and anti-Russian.

© 2011 Blackwell Publishing Ltd.

Results

Again, we examined both the number of trials to habituate (Figure 1a) and the overall time to habituate (Figure 1b). The number of habituation trials exhibited by infants presented with the learnable grammar ranged from 8 to 44 trials with a mean of 20.53 ($SD = 10.2$), while the number in the unlearnable condition ranged from 8 to 20 habituation trials with a mean of 13.6 ($SD = 5.19$). A *t*-test showed this difference to be significant ($t(28) = 2.34$, $p < .05$, two-tailed). With respect to overall time to habituate, infants exposed to the learnable grammar showed a range of times between 72 and 239 sec. with a mean of 115.36 ($SD = 56.8$), while infants in the unlearnable condition showed a range of times between 32 and 97 sec. with a mean of 64.3 ($SD = 21.77$). A *t*-test also showed the difference in total time to habituate to be significant ($t(28) = 3.25$, $p < .01$, two-tailed). Thus, as in Experiment 1, our main prediction was borne out for both dependent measures. Note that, because the patterns of results, and particularly the trials to habituate measure, were so similar between Experiments 1 and 2, the potential confound in the unlearnable stimuli of Experiment 2 is probably not driving the results.

Infants' longer listening times for learnable stimuli might be reflected in at least three trial-by-trial patterns, with each pattern potentially revealing something about how infants were processing the stimuli. One of these patterns would be for infants exposed to both learnable and unlearnable stimuli to show a steady decline in listening time over trials, but with a shallower slope for infants exposed to the learnable stimuli. Such a pattern might suggest that the learnable stimuli were only slightly more interesting to infants than the unlearnable stimuli. Another possibility would be for infants exposed to learnable stimuli to show little decline across earlier blocks of trials, followed by a rapid decline over later blocks. Such a pattern might suggest that infants attended to the learnable stimuli until they had extracted the relevant generalization, then lost interest. Finally, infants exposed to learnable stimuli might show a more jagged pattern of listening times, rising on some, falling on others, in contrast to a more steady decline for infants exposed to unlearnable stimuli. Such a pattern might suggest that infants are making new discoveries about the learnable input as they process it over time. All infants contributed a listening time for the first eight trials; therefore, we examined these trials to determine which of the three patterns of listening times better fit the data (see Figure 2). Figure 2 is most consistent with the third possible pattern. An analysis of the number of times that each infant showed a greater looking time on trial T than on trial T-1 revealed that infants presented with either version of a learnable grammar showed significantly more increases in listening time than infants presented with either version of the unlearnable grammar (learnable grammar $M = 3.47$, $SD = 0.86$; unlearnable
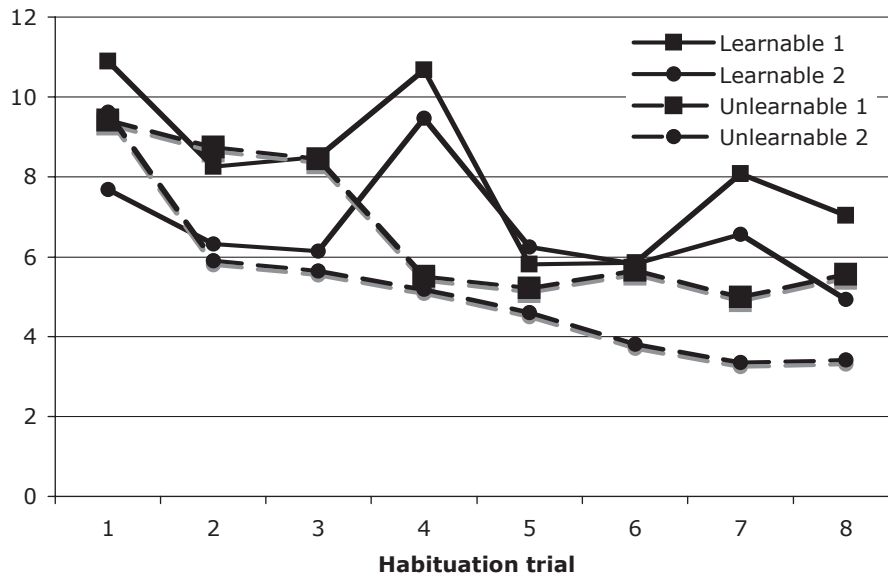
Figure 2 *Trial-by-trial listening times and standard errors for the first eight habituation trials in the first and second experiments.*

grammar $M = 3.0$, $SD = .87$, $t(58) = 2.09$, $p < .05$, two-tailed). We return to the implications of this analysis in the general discussion.

General discussion

Data from the two experiments show remarkable parallelism. In both experiments, infants stopped attending to the information source significantly sooner for the unlearnable than the learnable stimuli. This difference in attention was reflected both in the number of habituation trials and in the total time to habituate. Although the data are consistent with other studies of habituation, this is the first study in which trials to habituate and time to habituate have been examined for input that can be independently classified as learnable or unlearnable. Because the unlearnable condition of the second experiment was constructed from the two learnable conditions, it is unlikely that infants' differential listening to learnable vs. unlearnable grammars is due to low level stimulus differences. Rather, the most likely explanation for the pattern of data observed across the two experiments is that human learners under the age of 2 years attend differentially to learnable vs. unlearnable linguistic grammars.

Our proposal linking stimulus learnability and infant attention is as follows: Infants who were exposed to the unlearnable stimuli somehow detected that they were not making progress toward drawing any generalization from those stimuli. In contrast, infants who were exposed to the learnable stimuli somehow detected that they were making progress toward drawing a generalization, although they had not fully made the relevant generalization yet. The idea that infants are able to discern whether they are making progress toward learning

further suggests that infants have available to them an implicit metric of their own learning. What is the evidence that infants can detect that they are making progress? One source of support comes from an estimate about how long it actually takes infants to discover the appropriate generalization for the stimuli in Table 1. Earlier studies that tested infants' learning of Russian categories presented them with 2 min. of familiarization stimuli (the non-bolded stimuli in Table 1 or 2) followed by three blocks of test trials (the bolded stimuli and their ungrammatical counterparts; Gerken *et al.*, 2005). Only infants presented with the familiarization stimuli from Table 1 showed learning and only in the third block of test trials (after 2.5 min. or 150 sec. of stimulus exposure). Thus, it takes infants exposed to the stimuli in Table 1 a minimum of about 150 sec. to make the relevant generalization. Interestingly, this figure is very close to the average time to habituate in Experiment 1 (149 sec.). But if it takes an average of 149–150 sec. to make the relevant generalization, what in the input keeps infants listening for those 149–150 sec., given that infants exposed to the unlearnable input in Experiments 1 and 2 listened for an average of only 64 and 84 sec., respectively?

The pattern of data in Figure 2 and the analysis showing that infants who were exposed to learnable stimuli showed more rises in listening times over trials suggests that these learners continued to discover new potential generalizations either in the input itself or as they performed calculations on the input. We can identify at least two bases for such discoveries. One is more specific to the particular learning problem we chose. Braine suggested that learning grammars such as the one in Table 1 occurs in stages, such that learners first notice the existence of the unique category marker (*k* and *tel* in the experiments presented here). Learners

then, according to Braine, begin to associate the category marker with the other cues to categories (the case endings *oj* and *u* or *ya* and *yem*). Braine's characterization suggests that the particular problem we chose comprised sub-problems, with solving the first sub-problem perhaps keeping learners interested long enough for them to then solve the second sub-problem. This view of the data suggests that we should only expect to find differences in interest for learnable vs. unlearnable stimuli if the former comprise sub-problems, making the results of our experiments interesting but not very general. However, there is at least one reason to doubt this view: It would seem to predict one or two rises in interest for learnable stimuli over the course of all habituation trials. In contrast, infants who were exposed to learnable trials showed rises on approximately half of the first eight trials. That is, the pattern of rises and falls in interest does not seem to mirror a sub-problem structure in the stimuli. An alternative to the sub-problem view of infants' attention is that the rises and falls in interest we see in Figure 2 and associated analysis might reflect each infant's moment-to-moment successes and failures in finding local patterns in the input. This view of the data predicts that we will see differences in interest for learnable vs. unlearnable input across a variety of domains and learning problems. Importantly, it suggests that learners' interest is driven by a metric of their own learning, not by the stimulus structure alone.

Although we can currently say little about the nature of this learning metric, it is interesting to note that the compressibility notion of learnable stimuli might ultimately be drawn upon as a possible metric: If learning entails reducing the number of bits required to represent the input, then as long as the number of bits is decreasing, the learning problem remains interesting. When the number of bits fails to decrease or no longer decreases, learners might abandon the problem as either unlearnable or already learned. Although deriving a behavioural measure to test this notion might be difficult, it is certainly testable within various machine learning frameworks.

Regardless of the underlying cognitive or neurological status of the proposed metric, its apparent use by infants has at least four implications. First, infants' abilities might be viewed as an implicit parallel to the explicit metacognitive abilities observed in college students (Metcalfe & Kornell, 2005), in which adults devote more study time to learning problems on which they believe they are making progress, suggesting that implicit precursors to explicit metacognition are available from a very early age (Balcomb & Gerken, 2008). Second, progress in learning may be self-reinforcing, explaining why infants and young children commit mental resources to making complex generalizations in the laboratory and in the world without any obvious external reinforcement. A corollary to this second point is that some disorders of learning may reflect abnormalities of the system of self-reinforcement, not the generalization mechanism itself.

Third, computational models of learning in general, and of language learning in particular, might benefit from incorporating a metric of learning progress. Finally, the data provide a new insight into why human infants and children are such impressive learners – they devote their attention to aspects of the world in which valid generalizations can be made and avoid labouring in vain over unyielding input.

Acknowledgements

References

Balcomb, F.K., & Gerken, L.A. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, **11**, 750–760.

Braine, M.D.S. (1963a). On learning the grammatical order of words. *Psychological Review*, **70** (4), 323–348.

Braine, M.D.S. (1963b). The ontogeny of English phrase structure. *Language*, **39**, 1–14.

Cohen, L.B., Atkinson, D.J., & Chaput, H.H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies* (Version 1.0). Austin, TX: University of Texas.

Cultice, J., Somerville, S., & Wellman, H. (1983). Preschoolers' memory monitoring: feeling-of-knowing judgments. *Child Development*, **54** (6), 1480–1486.

Flavell, J., Green, F., & Flavell, E. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development*, **1** (1), 97–112.

Gerken, L.A., Wilson, R., Gomez, R., & Nurmsoo, E. (2009). The relation between linguistic analogies and lexical categories. In J.P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 101–117). Oxford: Oxford University Press.

Gerken, L.A., Wilson, R., & Lewis, W. (2005). 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language*, **32**, 249–268.

Gómez, R.L., & Gerken, L.A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, **70** (2), 109–135.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal understanding: the evolution, function and phenomenology of the theory-formation system. In F. Keil & R. Wilson (Eds.), *Cognition and explanation* (pp. 299–323). Cambridge, MA: MIT Press.

Hunter, M., & Ames, E. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, **5**, 69–95.

Hunter, M., Ames, E., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, **19** (3), 338–352.

Johnson, G. (1995). *Fire in the mind: Science, faith, and the search for order*. New York: Random House.

Lany, J.A., Gómez, R.L., & Gerken, L.A. (2007). The role of prior experience in language acquisition. *Cognitive Science: A Multidisciplinary Journal*, **31** (3), 481–507.

Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgments. *International Journal of Behavioral Development*, **26** (4), 327–333.

Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, **283**, 77–80.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, **18**, 159–163.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, **52**, 463–477.

Mintz, T.H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, **30** (5), 678–686.

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition*, **94**, 215–240.

Nelson, T.O., & Leonesio, R.J. (1988). Allocation of self-paced study time and the 'labor-in-vain effect'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14** (4), 676–686.

Quinn, P.C., & Bhatt, R.S. (2005). Learning perceptual organization in infancy. *Psychological Science*, **16** (7), 511–515.

Schulz, L., & Bonawitz, E.B. (2007). Serious fun: preschoolers play more when evidence is confounded. *Developmental Psychology*, **43** (4), 1045–1050.

Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press.

Soderstrom, M., & Morgan, J.L. (2007). Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech. *Developmental Science*, **10**, 641–653.

Smith, K.H. (1966). Grammatical intrusions in the recall of structured letter pairs: mediated transfer or position learning? *Journal of Experimental Psychology*, **72**, 580–588.