Regularity extraction across species:
associative learning mechanisms shared by human and non-human primates

Arnaud Rey, Laure Minier, Raphaëlle Malassis,
Louisa Bogaerts, and Joël Fagot

Laboratoire de Psychologie Cognitive & Brain and Language Research Institute
CNRS & Aix-Marseille Université, Marseille, France

Running Head: REGULARITY LEARNING

* Corresponding author:

Arnaud Rey
Laboratoire de Psychologie Cognitive
CNRS & Aix-Marseille Université
3, place Victor Hugo - Case D
13331 Marseille Cedex 03 – France
E-mail: arnaud.rey@univ-amu.fr

Abstract

Extracting the regularities of our environment is a core cognitive ability in human and non-human primates. Comparative studies may provide information of strong heuristic value to constrain the elaboration of computational models of regularity learning. The current study illustrates this point by testing human and non-human primates (Guinea baboons, *Papio papio*) with the same experimental paradigm, using a novel online learning measure. For *local* co-occurrence regularities, we found similar patterns of regularity extraction in baboons and humans. However, only humans extracted the more *global* sequence structure. It is proposed that only the first result that is common to both species should be used to constrain models of regularity learning. The second result indicates that the extraction of global regularities cannot be accounted for by mere associative learning mechanisms and suggests that humans probably benefit from their language recoding abilities for extracting these regularities. We propose to use a comparative approach to address a series of remaining theoretical questions, which will contribute to the development of a general theory of regularity learning.

Keywords: regularity extraction; statistical learning; implicit learning; associative learning; verbal recoding

Statistical learning – the ability to extract and encode environmental regularities – appears as a key feature of human cognitive systems and has been a domain of intensive research over the last 20 years (see Frost, Armstrong, Siegelman, & Christiansen, 2015, and Thiessen, Kronstein & Hufnagle, 2013, for reviews, but also e.g., Estes, 1950; Restle, 1970, for earlier pioneer work). It is thought to play a key role in the segmentation of continuous input, prediction, discrimination and categorization, shaping our basics representations of the environment (Frost et al., 2015).

Historically, before the ‘statistical learning’ wave, similar issues were addressed in the field of ‘implicit learning’. The latter term captures the observation that some forms of learning occur after the mere repeated exposure to regularities (or rule systems, as in artificial grammar learning experiments, e.g., Reber, 1967), with no intention of learning and no clear awareness of what has been learned (Cleeremans, Destrebecqz, & Boyer, 1998; Shanks, 2005). Some authors, such as Conway and Christiansen (2006), have proposed to merge the two terms (i.e., ‘implicit statistical learning’) to cover the general learning ability to extract and encode different forms of environmental regularities.

Others, like Perruchet (2005; see also Perruchet & Poulin-Charronnat, 2012), have argued that ‘statistical learning’ is nothing more than a new terminology corresponding, in fact, to the old domain of associative learning (e.g., Hebb, 1961; Mitchell & Le Pelley, 2010; Shanks, 1995). In this view, the extraction of regularities is understood as the formation of new larger units also called ‘chunks’. This process is supported by associative learning mechanisms, which itself is considered a by-product of the concurrent attentional processing of multiple elements (or small units). Learning can also be qualified as ‘implicit’ as it doesn’t require awareness of what is learnt nor the intention to learn (see also Perruchet & Pacton, 2006, Perruchet & Vinter, 2002). In that perspective, implicit statistical learning and associative learning are considered as two sides of the same coin, the former having a broader scope (including issues related to various aspects of language processing) compared to the latter that was more concerned by the formation of simple associations or by operant conditioning behaviors.

Basic associative (or implicit statistical) learning mechanisms are certainly shared by human and non-human primates (e.g., Wilson, Marslen-Wilson, & Petkov, 2017), and possibly by all species equipped with a nervous system (e.g., chicks: Santolin, Rosa-Salva, Regolin, & Vallortigara, 2016; songbirds: Chen, van Rossum, & Ten Cate, 2015; rodents: Toro, Nespor & Gervain, 2016). This raises the question which aspects of the human ability for regularity extraction can be related to those in non-human primates. The present study aims to address this question by a direct behavioral comparison between humans and Guinea baboons (*Papio papio*) in the same experimental paradigm. This paradigm provides an online measure of regularity extraction at different levels of integration. We argue that the use of an online learning measure on the one hand and the study of non-human primates on the other hand form a valuable approach to overcome two difficulties that characterize the existing literature.

The first difficulty is that the temporal dynamics of regularity extraction is frequently not considered. The seminal study of Saffran, Aslin, and Newport (1996) is often considered as the critical starting point in the study of statistical learning, and several theoretical accounts and computational models of this phenomenon have been proposed (e.g., Frank, Goldwater, Griffiths, & Tenenbaum, 2010; French, Addyman, & Mareschal, 2011; Perruchet & Vinter, 1998; Pothos, 2007). The empirical evidence used to elaborate these models is mainly derived from studies using the same artificial language learning paradigm. In this task, participants are exposed to an artificial language composed of few polysyllabic nonsense words (e.g., 4) that are auditorily presented without any pause between words. After this exposure phase, participants are asked to indicate which one of two test sequences looks more familiar. The main dependent variable used to test models is therefore an offline measure that informs us about which kind of regularities can be learned but that does not provide any information regarding the temporal dynamics of learning (see e.g., Siegelman, Bogaerts, & Frost, 2017, for a detailed discussion of the problems of offline learning measures). In contrast to offline measures, online measures are obtained by recording the performance of participants during the exposure to the regularities and can therefore provide information on the temporal dynamics of regularity extraction. The serial response time task (Nissen & Bullemer, 1987) is one good example of such online measure. In this task, participants respond to sequences of stimuli that appear one-by-one at various locations on a computer screen. If the sequences include regularities (allowing for the prediction of upcoming locations), then participants speed their response times (RTs) if they manage to extract these regularities (see also Batterink, 2017; Misyak, Christiansen, & Tomblin, 2010; Siegelman, Bogaerts, Kronenfeld, & Frost, 2017, for other examples of online learning measures).

A second difficulty is paradoxically the strong focus on human (adult) participants. Indeed, during the exposure phase, participants may develop strategies or explicit learning procedures (like repeating internally some sequences of syllables) and their resulting performance is certainly influenced and biased by various forms of explicit recoding processes. Testing human infants without any advanced language and reasoning abilities (like the original Saffran et al. study did, followed by many others, e.g., Bulf, Johnson, & Valenza, 2011; Singh, Reznick, & Xuehua, 2012) could be a solution to overcome that difficulty. However, the study of infants also brings other procedural difficulties because the measures that can be recorded for this population (for example, the head-turn procedure) are indirect and much noisier (i.e., less reliable). Testing non-human primates allows us to study regularity extraction in the absence of any bias due to language experience and the related ability to verbally recode regularities.

Previous studies that adopted a comparative approach to study statistical learning were mainly concerned with discovering differences between human and non-human primates (e.g., Fitch & Hauser, 2004; Wang, Uhrig, Jarraya, & Dehaene, 2015). That logic is derived from a research tradition that searched for a critical distinctive feature that could distinguish animal communication systems from human language. For instance, Hauser, Chomsky, and Fitch (2002) put forward the ability to process and produce recursive structures as a fundamental distinctive feature that sets humans apart from other species. However, as argued by Pinker and Jackendoff (2005), there are certainly many other critical differences between human and non-human primates that can account for their distinct cognitive trajectories (see also, Fagot, Malassis, Medam, & Montant, in press).

Here we propose an alternative perspective stating that we can increase our understanding of the precise dynamics of regularity learning processes by looking at the *common* patterns produced by human and non-human primates in online experimental paradigms. As mentioned above, it is obvious that human adults can make use of both implicit associative learning mechanisms and explicit recoding strategies based on their language abilities to perform regularity extraction tasks. Non-human primates miss the language tools that humans can use to explicitly recode regularities. One can therefore safely assume that non-human primates mainly rely on associative learning mechanisms in these tasks, and that these mechanisms are shared with humans. Therefore, any common pattern of performance between human and non-human primates should reflect common fundamental properties of these learning mechanisms. Alternatively, any difference could be related to the use of language recoding abilities and should therefore be considered with caution for constraining the development of computational models.

We now illustrate this proposition by comparing the performance of human and non-human primates in a novel task that has been designed to combine the serial response time task and the artificial language paradigm (Franco & Destrebecqz, 2012). Human adults and baboons had to touch a red circle appearing successively at nine different locations of a computer screen. The dot to follow moved on the screen along paths containing statistical regularities. The use of this task allowed us (1) to test the two populations with the exact same experimental procedure, (2) to provide detailed information on the temporal dynamics of regularity extraction, and (3) to compare humans and baboons at different levels of integration, at the local level within the sequence (i.e., co-occurrences between individual elements), or at the level of the global sequence structure (see the Method section below for an operational definition of the distinction between local and more global regularities). Note that part of the primate data was reported in Minier, Fagot, and Rey (2015).

Methods

Participants were ten Guinea baboons (*Papio papio*, age range 3–15.5 years) from the CNRS primate facility in Rousset (France) and 5 human adults (age range 20–24 years) from Aix-Marseille University.

Baboons live within a larger group of 26 individuals, within a 700 m2 outdoor enclosure and had a permanent access to ten Automated Learning Devices for Monkeys (ALDM, for a detailed description, see Fagot & Bonté, 2010; Fagot & Paleressompoulle, 2009) equipped with a 19-inch touch screen and a food dispenser. The main feature of ALDM equipment is that a radio frequency identification reader (RFID) identifies each baboon via a microchip implanted in each arm. The baboons can therefore participate to the research at will, without being captured, as the test programs recognize them automatically. All baboons had previously participated to numerous computerized experiments using the ALDM test systems. The experiment was controlled by the E-Prime software (Version 2.0, Psychology Software Tools, Pittsburgh).

A trial began with the presentation of a fixation cross at the bottom of a touch screen. The cross stayed on the screen till it was pressed. After pressing on it, the fixation-cross disappeared and immediately 9 crosses representing locations were displayed on the screen, one of them being replaced by the target, a red circle. When the target was touched, it disappeared and was replaced by the cross. The next cross/location in the sequence was then replaced by the red circle until the end of the sequence. Reward (a small pellet) was provided at the end of a sequence of nine touches.

If participants touched an inappropriate location (incorrect trial), or failed to touch the screen within 5,000ms after the red circle's appearance (aborted trial), a green screen was displayed for 3000ms (baboons) / 1000ms (humans) as a marker of failure. Aborted trials were not counted as trials and were therefore presented again, while incorrect trials were not. The time elapsed between the appearance of the red circle and the participant's touch on this circle was recorded as the response time (RT). To learn the task, baboons initially received trials that were rewarded after one touch, after which the number of touches in a trial was progressively increased up to nine. Humans were simply instructed to touch the red circle as quickly as possible. They were also informed that each trial began by the presentation of the yellow cross at the bottom of the screen followed by a sequence of nine touches on the moving red circle.

To equate for the motor difficulty of the sequences to be produced, each participant was first tested on 5 blocks (baboons) / 4 blocks (humans) of 200 random trials (called "Random I phase"), each composed of a random ordering of nine locations, without immediate repetition. A baseline measure for all possible transitions from one location to another was computed by calculating mean RTs for each transition at the group level (i.e., the entire group of baboons on the one hand and of human participants on the other hand).

After this Random I phase, participants were exposed to trials including statistical regularities (hereafter, “Pattern phase”). Three independent regularities, each composed of three fixed locations were constructed (R_1 = ‘$A_1B_1C_1$’; R_2 = ‘$A_2B_2C_2$’; R_3 = ‘$A_3B_3C_3$’). To study the extraction dynamic of these regularities, we carefully selected these three patterns so that the mean RTs for the initial transition (‘A_iB_i’) and final transition (‘B_iC_i’) would not be statistically different considering the baseline measurements obtained for these transitions during the random phase. By matching these *local* adjacent dependencies at the onset of the Pattern phase, it was possible to study the fine-grained evolution of these RTs that provide a direct behavioral index of regularity extraction.

Moreover, to study the extraction dynamic at a more complex and *global* level, the three regular patterns appeared in each sequence of nine touches, increasing the predictability of each pattern in the sequence (1/3, ½, 1). For example, if participants are exposed to the regular patterns R3 followed by R1, then the last regular pattern will be R2. If they manage to extract such a complex regularity, we should then observe a decrease in RTs for the first element of the last displayed regular pattern (in our example, for A_2).

During the Pattern phase, participants performed 10 (baboons) / 7 (humans) blocks of 200 trials each, one trial being composed of a random combination of the three regular patterns. Finally, participants performed a last series of 5 (baboons) / 1 (humans) random blocks (“Random II phase”) that had the same structure as the Random I phase (see Figure 1).

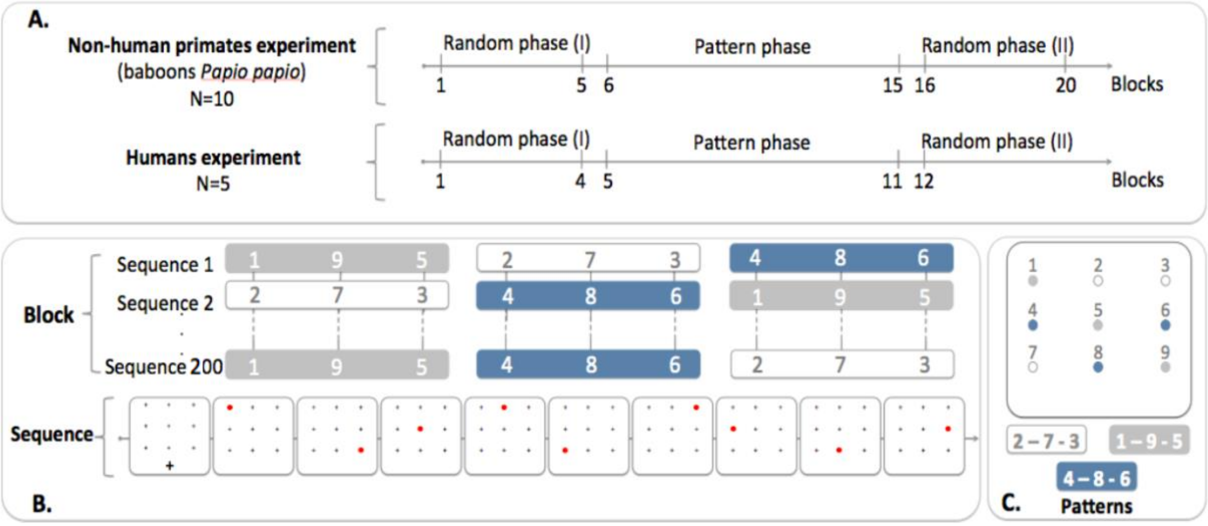

Figure 1: Experimental procedure for baboons and humans across the whole experiment (A). Experimental procedure at the block- and trial level (B). Patterns of locations on the touch screen used with humans (C).

Results

Incorrect trials were removed from the dataset (7.6% for baboons and 5.9% for humans), as well as RTs greater than two standard deviations from the mean (computed for each subject and each block, 4.2% for baboons and 3.6% for humans).

The first analysis concerned regularity extraction at the local level. Minier et al. (2015) reported an advantage of the second transition (i.e., BC) over the first one (i.e., AB) in baboons, with RTs on C decreasing faster than RTs on B. We therefore ran the same 2 (Transitions) * 7 (Block) repeated measures ANOVA on the mean RTs of humans. Mauchly's test indicated a violation of the sphericity assumption for the block factor ($\chi^2(44)=71.7$, $p<0.05$), so Greenhouse-Geisser estimates ($\varepsilon=0.11$) were used to correct for degrees of freedom. We found a significant effect of Transition ($F(1,9)=14.9$, $p<0.01$, $\eta^2_p=0.62$), first transition RTs being slower than those of second transitions, and a significant effect of Block ($F(2.88, 25.9)= 36.5$, $p<0.001$, $\eta^2_p=0.80$) as the response latencies decreased through learning. Finally, the Transition * Block interaction was also significant ($F(9,81)=4.9$, $p<0.001$,$\eta^2_p=0.33$). Contrasts between first and second transitions for each block revealed a significant difference in blocks 1 to 8, but not in blocks 9 and 10. As shown in Figure 2, the human results match those of baboons with a faster decrease of RTs on C than on B.

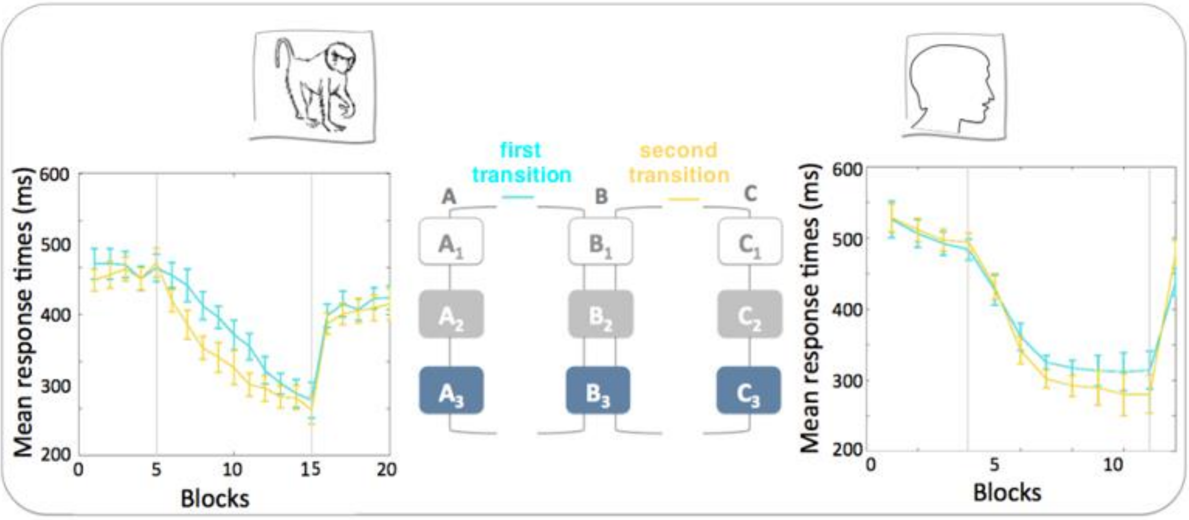

Figure 2: Evolution of mean RTs for the first (AB) and second (BC) transitions over blocks for baboons (on the left) and humans (on the right).

The second analysis concerned regularity extraction at a more global level, specifically, the predictability of the third pattern in the sequence given the first and second. As shown in Figure 3 (upper panel), the predictability within regular patterns is always equal to 1. Indeed, in each regular sequence, A_i was always followed by B_i that was always followed by C_i. However, the predictability of A varied across patterns depending on their position in the sequence of 9 touches. The first A of the first ABC regular pattern could occur with a probability of 1/3, the second A with a probability of 1/2, and the last A with a probability of 1. Of course, these probabilities depend on the ability of participants to extract the global statistical structure of the experiment.

We thus considered RTs for the three patterns on the three different positions in the sequence. Because humans and baboons did not receive the same number of blocks, we ran two different ANOVAs, both including Place in the pattern (first = A, second = B or third =

C), Blocks (1 to 10 for baboons and 1 to 7 for humans) and Position of the patterns in the sequence (first, second or third) as within-subject factors.

For baboons, Mauchly's test indicated a violation of the assumption of sphericity for the Block factor ($\chi2$ (44)=67.5, $p<0.05$) as well as for the Position factor ($\chi2$ (2)=6.3, $p<0.05$), so Greenhouse-Geisser estimates (ε=0,31 and ε= 0.62, respectively) were used to correct for degrees of freedom. We found a significant effect of Position (F(1.23, 11.2)=8,0, p<0,013, η^2_p=0,47), and a significant effect of Block (F(2.8, 25.2)= 3.7, p<0.001, η^2_p=0.29). The interaction between Block and Position were found to be significant as well (F(18,162)=1.8, p=0.026, η^2_p=0.17). Helmert contrasts showed that the effect of Position was between the first position and the last two (F(1,9)=8.9, p=0.02, η^2_p = 0.5) but not between the second position and the third (F(1,9)=1.74, p=0.22, η^2_p =0.16). Another ANOVA ran on the mean RTs of the first place of the patterns (i.e., A) with only the last learning block and Position as a within-participant factor with 3 levels (1st, 2nd and 3rd) showed a significant effect (F(2,18)=4.9, p<0.05, η^2_p =0,35). Helmert contrasts confirmed that this effect was significant between the first and two last positions (F(1,9)=6.2, p=0.04, η^2_p =0.40) but not between the second and the third positions (F(1,9)=0.67, p=0.43, η^2_p =0.07).

For humans, we found a significant effect of Position (F(2,8)=36,83, p<0,001, η^2_p=0,90), and a significant effect of Block (F(6, 24)= 7,0, p<0.001, η^2_p=0.64). The interaction between Position and Block was also significant (F(12,48)=13.57, p<0.001, η^2_p=0.77). Helmert comparisons showed that the effect of Position was between the first position and the two last (F(1,4)=51.5, p=0.02, η^2_p= 0.98), and between the second position and the third position (F(1,4)=8.03, p=0.047, η^2_p=0.67). A final ANOVA considered mean RTs of the first place in the patterns (i.e., A) in the last learning block only. Position was entered as within-participant factor with 3 levels (1st, 2nd and 3rd). This analysis showed a significant effect (F(2,8)=36.7, p<0.0001, η^2_p=0,90). Helmert contrasts indicated that this effect was significant between the first and two last positions (F(1,4)=68.0, $p<0.001$, $\eta^2_p=0.94$) and between the second and the third position as well (F(1,4)=14.3, $p=0.02$, $\eta^2_p=0.78$). The evolution of mean RTs for baboons and humans over each block, each position within the sequence, and each place within the regular pattern is provided in Figure 3 (lower panels). A clear difference between baboons and humans appears for the first element of the triplet when it occurred in Position 3 of the sequence (i.e., A_3): RTs decreased across learning blocks for humans while no decrease was observed for baboons.

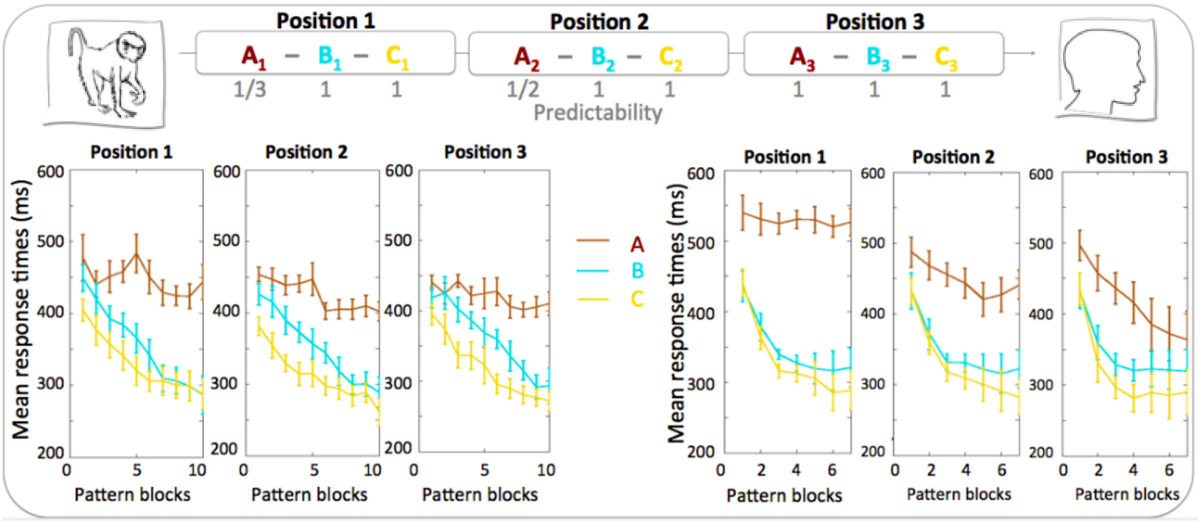

Figure 3: Upper central panel: Predictability of each location in the sequence of nine touches. Lower panels: Evolution of mean RTs over blocks (1-10 for baboons and 1-7 for humans), positions (1, 2, or 3) and places within the regular patterns (A, B, or C) for baboons (on the left) and humans (on the right).

Discussion

Two main findings were obtained in the present comparative study in which the performance of human adults and baboons were collected in the same regularity learning task. For regularities at the *local* level, human and non-human primates displayed a similar statistical learning dynamic: RTs decreased faster on the third location of a regular pattern (i.e., C) than on the second (i.e., B). At a more *global* level, we found different results for humans and baboons. In humans, RTs for the first element (i.e., A) of the third displayed regular pattern (i.e., the 7th element in the sequence, which should become predictable over the course of learning) decreased gradually across learning blocks while no such improvement was observed for baboons.

The first result, already reported for baboons by Minier et al. (2015), indicates that the two species had a similar dynamic of regularity extraction. In both species, RTs decreased faster on C than on B in ABC regular patterns, suggesting that it is a general property that should be captured by current computational models. As argued previously by Minier et al. (2015), this pattern of results is inconsistent with theoretical propositions resting on the notion of transitional probabilities between adjacent elements. Indeed, in our task, the transitional probabilities from A to B and from B to C are all equal to one and are therefore insufficient to account for the present data. Similarly, the TRACX model by French et al. (2011) predicts that RTs should decrease faster on B compared to C given that the model is based on a left-to-right acquisition of statistical regularities. Only the SRN model (Elman, 1990) predicts the learning advantage of C over B. SRN assumes that learning is based on the development of associations between the successive elements of the input that are frequently and consistently encountered together. SRN encodes also the contextual information that precedes the successive elements, and therefore C benefits from richer contextual information (i.e., the preceding systematic sequence AB) while B benefits only from the systematic occurrence of A.

Our second result indicates that only humans capture the more complex regularity related to the predictability of the third regular pattern in the larger sequence. In that situation, following the logic that we proposed, the absence of a common behavioral pattern of results between human and non-human primates suggests that this kind of regularity cannot be extracted by mere associative learning mechanisms. Computational models are therefore not expected to account for this phenomenon that may require language recoding abilities.

Following a different line of reasoning, one may argue that the present results point to a possible limitation of the learning capacities in non-human primates. The extraction of these global regularities might indeed require a more complex learning mechanism or a larger memory span for preceding elements. One would need however to specify what is the precise nature of this learning mechanism and what would be the role of short-term memory span in facilitating the extraction of such global regularities.

Conclusions and avenues for future research

The comparative approach taken in the current paper suggests that the mechanisms underlying regularity extraction can be further understood by looking at patterns of performance that are common to human and non-human primates. Based on the present results we suggest that shared associative learning mechanisms can account for the extraction of *local* regularities (i.e., co-occurrences) and that a general theory of regularity learning could benefit from the large existing literature on associative learning (Mitchell & Le Pelley, 2010; Shanks, 1995). Moreover, the temporal dynamics of co-occurrence learning we observed in both species suggests a learning advantage for the final stimulus of a repeated (triplet) pattern, providing an important empirical constraint for models of regularity extraction. Sensitivity to *global* regularities at the level of the entire sequence was observed for humans only and might call for a theory that accounts for the interaction between basic associative learning mechanisms and verbal recoding strategies.

Several important issues remain to be addressed to reach a general theory of associative learning mechanisms and their role for regularity extraction in different situations. We propose that the use a comparative approach can be a fruitful avenue for future research to address a series of remaining theoretical questions, which we outline in more detail below.

A first series of questions concerns the role of the *timing* of co-occurrences in the sensory input. Suppose that A and B are two sources of information that co-occur systematically, A being always followed by B. Assuming no interfering information between A and B, how does the dynamic of regularity extraction change as a function of the delay between the presentation of A and the presentation of B? How many times A-B has to be repeated in order to observe an extraction of the co-occurrence? What is the role of the delay between two repetitions of A-B?

A second series of questions is related to the *sources of information surrounding* the A-B co-occurrence to be extracted. A general theory of regularity learning should be able to specify the role of factors such as the number and nature of interfering information between two repetitions of A-B, the role of the similarity between the interfering information and each element of the co-occurrence (A and B), the role of forward and backward inconsistencies (i.e., the role of the statistical co-occurrence of A with another source of information, e.g., C, and the interference of A-C on the extraction of A-B or, respectively, the co-occurrence of B with a preceding source of information, e.g., D, and the interference of D-B on the extraction of A-B), and finally, the effect of the presence of other co-occurrences (e.g., like E-F) appearing between two repetitions of A-B.

A third series of questions concerns the *number of co-occurring elements*. What is the regularity extraction dynamic when A-B is systematically followed by another information C (an issue that has started to be addressed in the present study)? Similarly, what will happen if we increase the chain of co-occurring elements? For ABCD sequences, for example, will we observe an even faster decrease of RTs on D compare to C (because of the richer context) or is the role of contextual information limited to a restricted time-window? And for longer sequences, what is the effect of repeating a previously encountered element (e.g., like in the sequence ABCAD)?

A fourth set of questions is related to the *learning of non-adjacent co-occurrences* (Gómez, 2002 ; Pacton & Perruchet, 2008 ; Pacton, Sobaco, & Perruchet, 2015). Under which conditions can we extract the co-occurrence of A and B when other sources of information are placed between A and B? Clearly, if co-occurrence learning is supported by associative learning mechanisms, the presence of interfering information between A and B will strongly decrease the probability to extract this regularity. However, baboons have shown some ability to process long-distance dependencies (Rey, Perruchet, & Fagot, 2012) indicating that the learning of non-adjacent regularities should be accounted for by current computational models of regularity learning.

To conclude, we have proposed an alternative way of using comparative studies to better comprehend the fundamental mechanisms supporting regularity extraction. Rather than focusing on the differences between species, we suggest focusing on common patterns of performances to determine the empirical evidence that should constrain the elaboration of computational models. By employing online measures, we can generate fine-grained behavioral results that can inform models about the precise dynamics of regularity extraction mechanisms. We have suggested that this approach can be quite powerful in testing the predictions of computational models. Finally, we have outlined a list of questions that would need to be addressed to reach a deeper understanding of associative learning mechanisms in the context of regularity extraction.

Acknowledgments

This work, carried out within the Labex BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). This research was also supported by the Premilang2 ANR-project (#ANR-13-BSH2-0002) and the Chunked ANR-project (#ANR-17-CE28-0013-02).

References

Batterink, L. J. (2017). Rapid Statistical Learning Supporting Word Extraction From Continuous Speech. *Psychological Science*, 28, 921-928. doi:10.1177/0956797617698226

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Chen, J., van Rossum, D., & Ten Cate, C. (2015). Artificial grammar learning in zebra finches and human adults: XYX versus XXY. *Animal Cognition*, *18*(1), 151-164. doi:10.1126/science.284.5416.875a

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406-416. doi:10.1016/S1364-6613(98)01232-7

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88 (1), 2–19.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179-211.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*(2), 94-107. doi :10.1037/h0058559

Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behavior Research Methods*, *42*(2), 507-516.

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods*, *41*(2), 396-404.

Fagot, J., Malassis, R., Medam, T. & Montant, M. (in press). Comparing human and nonhuman animal performance on domain-general functions: towards a multiple bottleneck scenario of language evolution. in *Origins of human language: continuities and splits with nonhuman primates.* Editors: Louis-Jean Boë, Joël Fagot, Pascal Perrier, Jean-Luc Schwartz, Peter lang.

Fitch, W. T., and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*, 377-380.

Franco, A., & Destrebecqz, A. (2012). Chunking or not chunking? How do we find words in artificial language learning? *Advances in Cognitive Psychology*, *8*(2), 144.

Frank, M. C., Goldwater, S., Griffiths, T., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117*, 107-125.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A Recognition-Based Connectionist Framework for Sequence Segmentation and Chunk Extraction. *Psychological Review, 118*(4), 614-636.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125. doi:10.1016/j.tics.2014.12.010

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford, UK: Blackwell.

Minier. L., Fagot, J., & Rey, A. (2015). The temporal dynamics of regularity extraction in nonhuman primates. *Cognitive Science*. doi:10.1111/cogs.12279

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 31. doi.org: 10.3389/fpsyg.2010.00031

Mitchell, C. J., & Le Pelley, M. E. (Eds.). (2010). *Attention and associative learning: From brain to behaviour*. Oxford University Press, USA.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 80-96.

Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based account of adjacent and nonadjacent dependency learning valid ? *Acta Psychologica, 157*, 195-199.

Perruchet, P. (2005). Statistical approaches to language acquisition and the self-organizing consciousness: A reversal of perspective. *Psychological Research*, 69, 316-329.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233-238. doi:10.1016/j.tics.2006.03.006

Perruchet, P., & Poulin-Charronnat, B. (2012). Word segmentation: Trading the (new, but poor) concept of statistical computation for the (old, but richer) associative approach. In Rebuschat, P., & Williams, J. N. (Eds.), *Statistical learning and language acquisition* (pp. 119-143). Berlin: De Gruyter Mouton.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246-263.

Perruchet, P., &Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences*, 25, 297–388. doi:10.1017/S0140525X02000067

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it?. Cognition , 95 (2), 201-236.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*(2), 227-244. doi: 10.1037/0033-2909.133.2.227

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855-863. doi: 10.1016/S0022-5371(67)80149-X

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*(6), 481-495. Doi: 10.1037/h0029964

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (Papio Papio). *Cognition*, *123*(1), 180-184.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Santolin, C., Rosa-Salva, O., Regolin, L., & Vallortigara, G. (2016). Generalization of visual regularities in newly hatched chicks (Gallus gallus). *Animal Cognition*, *19*(5), 1007-1017. doi:10.1007/s10071-016-1005-2

Shanks, D. R. (1995). *The Psychology of Associative Learning*. Cambridge: Cambridge University Press.

Shanks, D. R. (2005). Implicit learning. In: Lamberts, K., and Goldstone, R., (eds.) *Handbook of Cognition*, pp. 202–220. London: Sage.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior research methods, 49*(2), 418-432. doi:10.3758/s13428-016-0719-z

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2017). Redefining “Learning” in Statistical Learning: What Does an Online Measure Reveal About the Assimilation of Visual Regularities? *Cognitive Science*. doi:10.1111/cogs.12556

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental Science, 15*(4), 482-495. doi: 10.1111/j.1467-7687.2012.01141.x

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological bulletin*, *139*(4), 792-814. doi:10.1037/a0030801

Toro, J. M., Nespor, M., & Gervain, J. (2016). Frequency-based organization of speech sequences in a nonhuman animal. *Cognition*, *146*, 1-7. doi:10.1016/j.cognition.2015.09.006

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Current Biology, 25*, 1966-1974.

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends in Neurosciences*, 40, 72-82. doi:10.1016/j.tins.2016.11.004