

836 **Supplementary Materials** for Wilson *et al.*, ‘Mixed-complexity artificial grammar learning in humans
837 and macaque monkeys: Evaluating learning strategies’

838

839 **Supplementary Figure 1: Eye-tracking experiments testing AG learning in adult human participants**

840 In the main article we report data from an eye-tracking experiment in macaques and from a two-
841 alternative forced-choice experiment in humans. While it would be ideal to test the humans and
842 macaques with the same methods (for further discussion of this point, see Petkov & Wilson, 2012),
843 different species might find it more natural to respond in different ways, making using the same
844 tasks difficult. Here, we report two eye-tracking experiments conducted in human participants. In
845 the first experiment 11 participants were not given any instructions regarding where to look during
846 the stimulus presentation. The results show that adult humans did not tend to look towards the
847 speaker presenting the auditory AG sequences. In the second experiment we asked 10 different
848 participants to try to localise the source of the testing sequences by looking towards it. In this
849 experiment, on average participants did look towards the speaker for longer, but again the looking
850 responses did not relate to whether the sequences were consistent with the AG or whether they
851 contained violations. The same participants were then tested with a two-alternative forced-choice
852 experiment like the one described in the article and showed evidence of having learned the artificial
853 grammar. Thus neither of these two eye-tracking experiments were able to measure any effects of
854 AG learning based on adult humans’ natural looking responses.

855 **Stimuli**

856 The human eye-tracking experiments used the same AG as the main experiments reported here. The
857 stimulus sequences and nonsense words were the same as those used in our previous study (see
858 Suppl. Table 1 and Wilson *et al.*, 2013).

859 **Supplementary Table 1.** Stimuli for human eye-tracking experiments. Table adapted from Wilson *et*
860 *al.*, (2013).

AG Element	Nonsense Word	Exposure Sequences	Testing Sequences	Condition
A	‘klor’	ACF	ACGFC	Correct
C	‘biff’	ACFC	ADCFCG	Correct
D	‘jux’	ACGF	ACFCG	Correct
F	‘cav’	ACGFC	ADCGFC	Correct
G	‘dupp’	ACGFCG	AFGCD	Violation
		ADCF	AFCDGC	Violation
		ADCFC	FADGC	Violation
		ADCFCG	DCAFGC	Violation
		ADCGF		

861

862

863 **Human eye-tracking experiment 1: Free looking**

864 In the first human experiment we tested 11 participants (1 male, 10 female; age range 18-36 years).
865 The participants were initially exposed to the testing sequences binaurally for 5 minutes, without
866 explicit instruction about what they were listening to. As in the macaque experiment, in the
867 subsequent testing phase, the participants were presented with a fixation spot for 2 seconds, before
868 the spot disappeared and a stimulus was presented from one of two audio speakers located at $\pm 30^\circ$
869 visual angle concealed behind a black curtain. The participants were asked to fixate on the fixation
870 spot while it was present, but were given no further instructions about where to look during the
871 stimulus presentation. Eye-tracking data were recorded with the participants in an eye-tracking head
872 frame using the same infra-red eye-tracking system that we used with the macaques (see article
873 Methods). Eye-tracking data were recorded from the beginning of the fixation period until 5 seconds
874 after the stimulus was presented (for a total of 7 seconds of eye-tracking data). Each participant took
875 part in 5 testing runs (each containing 4 presentations of each stimulus). Each testing phase was
876 separated by a re-familiarisation period during which the participant listened to the exposure
877 sequences for 3 minutes prior to the start of the next testing run.

878 First, we plotted the participants' mean eye position relative to the presenting audio speaker
879 (Suppl. Fig. 1A). The participants fixated on the spot accurately and on average, the participants did
880 tend to look towards the presenting audio speaker, however these looking responses to the sound
881 from the hidden audio speaker were weak. The mean eye position during the stimulus presentation
882 was less than 1° visual angle. This suggests that the participants typically kept looking towards the
883 centre of the screen, where the fixation spot had previously been displayed, rather than looking
884 towards the audio speakers. As in the macaque experiment, we analysed the duration of the
885 responses based on a response threshold calculated from the variability in the fixation period. The
886 participants fixated very tightly during the fixation period so this threshold was lower than for the
887 macaques; therefore small eye movements were considered as responses. However, there were no
888 differences between the responses to the consistent and violation sequences ($t_{10} = 0.336$, $p = 0.744$,
889 Suppl. Fig. 1B). We also calculated the mean horizontal eye position throughout the stimulus period.
890 Again, there was no significant difference in eye position between responses to consistent and
891 violation sequences ($t_{10} = 0.698$, $p = 0.501$, Suppl. Fig. 1C).

892 This experiment failed to provide any evidence for differences in responses, in either
893 direction, between consistent and violation AG sequences, possibly because the adult humans did
894 not make many looking responses towards the audio sequences. To address whether effects were
895 limited because the humans were insufficiently motivated to look toward the speakers, we
896 conducted a second experiment using identical stimuli and methods, in which we encouraged the
897 participants to look towards the presenting audio speaker to localise the sounds.

898

899 **Human eye-tracking experiment 2: Sound localisation**

900 In this second eye-tracking experiment we modified the instructions the participants received in
901 order to encourage them to make more looking responses. As in the first experiment the
902 participants were asked to fixate on the centrally located fixation spot. When the fixation spot
903 disappeared they were asked to localise the sound from the hidden speaker by looking towards it.

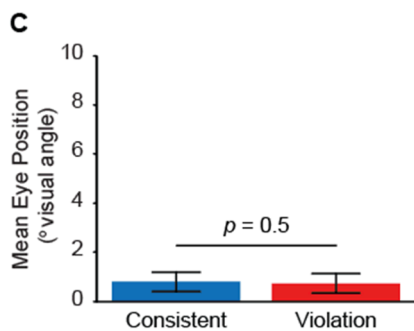
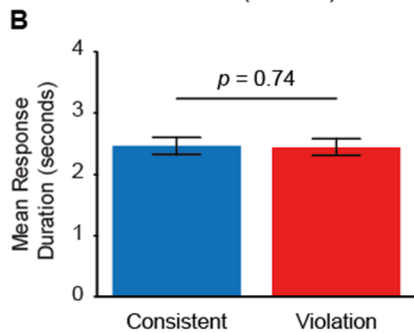
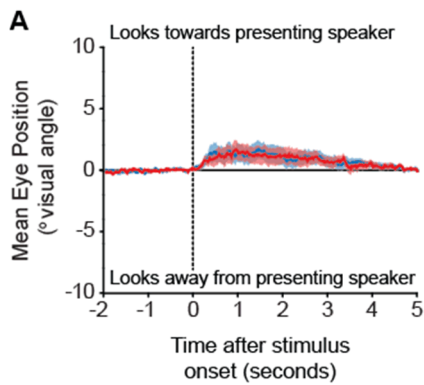
904 Other than these new instructions, the experiment and analyses were performed identically to the
905 previous experiment. 10 participants were tested in this experiment (4 male, 6 female; age range 18-
906 27 years).

907 In this experiment, the participants made much stronger looking responses towards the
908 presenting audio speaker than in the previous experiment (Suppl. Fig. 1D). However, the participants
909 did not respond for longer to either the consistent or violation sequences ($t_9 = 1.02$, $p = 0.332$, Suppl.
910 Fig. 1E). Furthermore, there was no significant difference in eye position during the presentation of
911 consistent vs violation sequences ($t_9 = 1.63$, $p = 0.14$, Suppl. Fig. 1F).

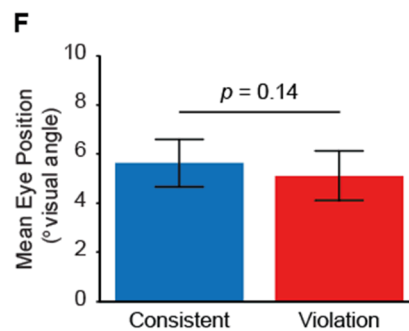
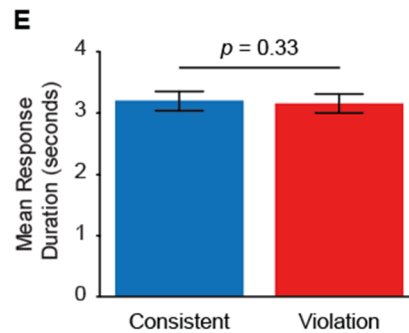
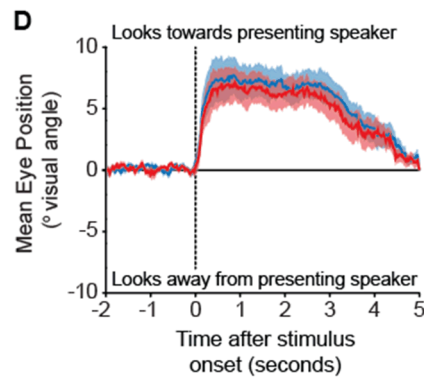
912 In summary, both of the eye-tracking experiments failed to provide any evidence from adult
913 human looking responses that artificial grammar learning has occurred. To determine whether this
914 result reflects a failure of these participants to learn the AG we also tested the participants of the
915 second experiment with a two-alternative forced-choice task, like the one used in the main
916 experiments here. Following the eye-tracking experiment each of the participants took part in a
917 single testing run where they were explicitly asked if the testing sequences followed the same
918 pattern as the exposure sequences or not. As in the main experiment, we found that the participants
919 gave significantly more “violation” responses to the violation sequences (violation vs consistent; $t_9 =$
920 2.42 , $p = 0.03$). This suggests that the participants were able to identify the violation sequences;
921 however, unlike in the macaques, we were unable to measure their AG learning abilities using eye-
922 tracking.

923

**Experiment 1:
Free looking**



**Experiment 2:
Sound localisation**



924

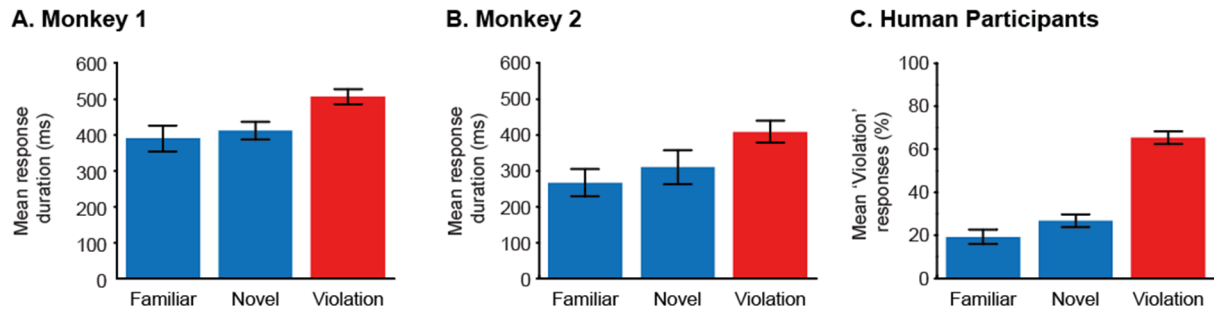
925 **Supplementary Figure 1. Human eye-tracking experiments.** (A-B) Mean (\pm SEM) eye position during
 926 the fixation period (from -2 to 0 seconds before stimulus presentation) and following the stimulus
 927 presentation (from 0-5 seconds). Responses to violation sequences are shown in red, those to
 928 consistent sequences are shown in blue. Positive eye positions represent looks towards the
 929 presenting speaker (whether the speaker was on the right or left of the monitor). (C-D) Mean
 930 durations (\pm SEM) of responses to the consistent and violation sequences, measuring the duration for
 931 which the eye position exceeded the threshold generated from variability in response in the baseline
 932 period (see Methods). (E-F) Mean (\pm SEM) eye position during the stimulus presentation period (0-5
 933 seconds in A-B).

934

935 **Supplementary Figure 2: Responses to familiar and novel consistent testing sequences, relative to**
936 **violation sequences, in the macaques and humans**

937 In the main analyses we noted that a significant difference was observed between the familiar and
938 novel consistent sequences in the human participants, but not in the macaques (see Results).
939 Evaluating the results more closely across the species suggests qualitatively similar patterns,
940 although the human, but not the macaque results, are statistically significant (see article Results).

941



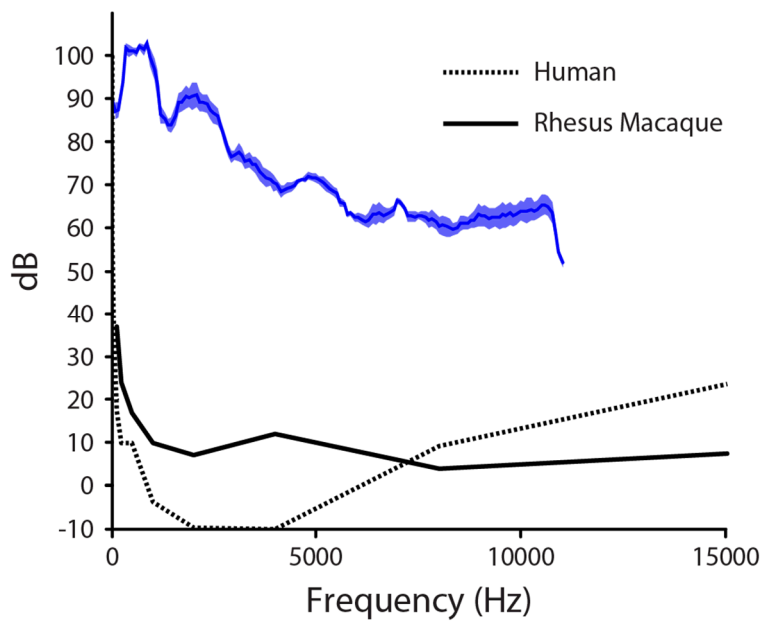
942

943 **Supplementary Figure 2. Responses to familiar and novel consistent testing sequences, relative to**
944 **violation sequences, in macaques and humans.** (A-B) Mean (\pm SEM) response duration to the
945 different testing conditions in the two macaques. (C) Mean (\pm SEM) percentage of 'Violation'
946 responses in the human participants. In both of the monkeys and in the human participants, we
947 observed stronger responses to the violation sequences than the consistent ones. In the human
948 participants there was also a significant difference between the familiar and novel consistent
949 sequences, which was not statistically significant in the macaques.

950

951 **Supplementary Figure 3: Nonsense word power spectra in relation to human and macaque**
952 **audiograms**

953 The nonsense word stimuli in this experiment were designed to be well within the hearing range of
954 both humans and macaques, which are very comparable based on published data. Suppl. Fig. 3
955 shows the average power-spectrum (\pm SEM) of the acoustic content in the 5 nonsense words. These
956 stimuli were calibrated at an RMS sound level of \sim 75dB. On this plot we also show published data on
957 human and macaque audiograms. The result shows that the stimuli were well within the audible
958 range of both species, and that at least in this range, the stimuli were well within the audible range
959 of the humans and macaques.
960



961

962 **Supplementary Figure 3. Nonsense word power spectra in relation to human and macaque**
963 **audiograms.** Blue filled area signifies the range (\pm SEM) of the acoustic power spectrum of the
964 nonsense word sounds (in dB re 20 μ N/m²). The black lines indicate the auditory threshold above
965 which sounds at that frequency and intensity are audible for humans (dotted line) and macaques
966 (solid line). The nonsense word stimuli fall well within the audible range of both the humans and
967 macaques. Human audiogram data from: Jackson *et al.*, (1999). Macaque audiogram data
968 summarised from: Pfingst *et al.*, (1975), Pfingst *et al.*, (1978), Lonsbury-Martin & Martin, (1981) and
969 Bennett *et al.*, (1983)

970
971
972
973

974 **Supplementary Text: Responses to the first presentation of each consistent testing sequence**

975 In order to balance the number of presentations of the 8 violation testing sequences it was
976 necessary to present each of the 4 consistent testing sequences twice each (Fig. 1). In order to assess
977 whether the results reported in the article would persist when only the first presentation of each
978 consistent sequence was considered, we reanalysed the data excluding the second presentation of
979 each consistent sequence. Although omitting half of the consistent sequences reduces the power of
980 the analyses, the results are comparable to those reported in the article.

981 An RM-ANOVA including the repeated measure of Condition (with two levels: consistent or
982 violation sequence) and the between subjects factor of Monkey (two levels) produced a strong main
983 effect of Condition ($F_{1,30} = 17.548, p < 0.001$). There was no interaction between Condition and
984 Monkey ($F_{1,30} = 0.303, p = 0.586$). These observations show that the effects reported in the main
985 experiment are also present when only the first presentation of each consistent sequence is
986 considered. Next, we performed an RM-ANOVA including the three levels of familiar, novel and
987 violation as a Condition factor. Again we saw a strong main effect of condition ($F_{2,30} = 9.01, p = 0.001$)
988 and no interaction between Condition and Monkey ($F_{2,30} = 0.441, p = 0.648$). Bonferroni corrected
989 post-hoc tests revealed a significant difference between familiar and violation sequences ($p = 0.009$)
990 and a statistical trend towards a significant difference between the novel and violation sequences (p
991 $= 0.1$). There was no significant difference between the familiar and novel sequences ($p = 1.0$).
992 Although these results are underpowered relative to those in the article, they support the same
993 conclusions. These results do not suggest that any differences between the consistent and violation
994 sequences can be attributed to attenuated looking responses to the consistent sequences caused by
995 the second repetition of these sequences.

996 Finally, we performed the same analyses on the human data. A paired samples t -test
997 revealed that participants gave significantly more 'violation' responses to the violation sequences
998 than to the first presentations of the consistent sequences ($t_{32} = 8.03, p < 0.001$). An RM-ANOVA
999 including the factor Condition (with three levels: familiar, novel and violation sequences) showed a
1000 strong main effect of Condition ($F_{2,64} = 47.2, p < 0.001$). Post-hoc tests revealed significant
1001 differences between the familiar and violation sequences ($p < 0.001$) and the novel and violation
1002 sequences ($p < 0.001$), but no difference between the familiar and novel consistent sequences ($p =$
1003 0.367). These results recapitulate those in the main article.

1004

1005

1006

1007 **Supplementary Text: Analyses of responses across testing runs**

1008 The monkeys participated in multiple testing runs in each testing session. To assess whether the
1009 monkeys might show diminished responses to the violation sequences after long sessions, we
1010 calculated the difference in response duration to the violation and consistent sequences for each
1011 testing run (violation – consistent) and correlated this with the number of testing runs the monkey
1012 had participated in that day. Neither monkey showed a significant correlation (M1: $r = 0.327$, $p =$
1013 0.216 ; M2: $r = 0.216$, $p = 0.572$), demonstrating that the violation effect did not diminish on days
1014 with many testing runs. Furthermore, there was no correlation between the difference in response
1015 durations and the total number of testing runs the animal had participated in across all sessions (M1:
1016 $r = -0.26$, $p = 0.922$; M2: $r = 0.121$, $p = 0.656$). These analyses demonstrate that the monkeys'
1017 responses did not vary with the amount of testing they had received.

1018 Additional analyses were performed to assess whether stronger responses to novel than
1019 familiar consistent sequences might have occurred in earlier testing runs but diminished with
1020 repeated testing. The difference between the animals' response durations to the novel and familiar
1021 sequences (novel – familiar) was calculated and correlated with the number of testing runs. There
1022 was no significant correlation between the responses to the novel relative to the familiar sequences
1023 and repeated testing in either animal (M1, $r = -0.148$, $p = 0.583$; M1, $r = -0.217$, $p = 0.418$). These
1024 results suggest that the monkeys' did not produce diminishing responses to the novel consistent
1025 sequences over multiple testing runs.

1026 **Supplementary Text: Analyses of human participants' reaction times.**

1027 Reaction times were recorded in the main human experiment. However, in order to ensure that the
1028 participants listened to the entire sequence before responding (rather than responding early and
1029 missing violations or important transitions later in the sequences), responses were only allowed
1030 after the entire sequence had been presented. Therefore, the reaction times do not clearly reflect
1031 how quickly participants recognised violations. Rather they reflect how quickly the human
1032 participants responded following the end of the sequence presentation.

1033
1034 The reaction time results were analysed in the same ways as the response data. A paired-
1035 sample t-test revealed no difference in RTs between the consistent and violation sequences ($t_{32} =$
1036 $0.423, p = 0.675$). An RM-ANOVA revealed no significant relationship between RTs and the number
1037 of rule violations in the sequences ($F_{3,128} = 1.731, p = 0.164$). There was also no correlation between
1038 RT and the mean TP of the sequences ($r = 0.029, p = 0.500$). Finally, there was no difference between
1039 violation sequences containing the 'ACF' violation relative to those that did not ($t_{32} = 0.436, p =$
1040 0.666).

1041

1042

1043 **Supplementary References**

- 1044 Bennett, C.L., Davis, R.T. & Miller, J.M. (1983) Demonstration of presbycusis across repeated
1045 measures in a nonhuman primate species. *Behavioral Neuroscience*, **97**, 602-607.
- 1046 Jackson, L.L., Heffner, R.S. & Heffner, H.E. (1999) Free-field audiogram of the Japanese macaque
1047 (*Macaca fuscata*). *Journal of the Acoustical Society of America*, **106**, 3017-3023.
- 1048 Lonsbury-Martin, B.L. & Martin, G.K. (1981) Effects of moderately intense sound on auditory
1049 sensitivity in rhesus monkeys: behavioral and neural observations. *J Neurophysiol*, **46**, 563-
1050 586.
- 1051 Petkov, C.I. & Wilson, B. (2012) On the pursuit of the brain network for proto-syntactic learning in
1052 non-human primates: conceptual issues and neurobiological hypotheses. *Philos Trans R Soc
1053 Lond B Biol Sci*, **367**, 2077-2088.
- 1054 Pfingst, B.E., Hienz, R. & Miller, J. (1975) Reaction-time procedure for measurement of hearing;
1055 Threshold functions. *Journal of the Acoustical Society of America*, **57**, 431-436.
- 1056 Pfingst, B.E., Laycock, J., Flammino, F. & Lonsbury-Martin, B.L. (1978) Pure tone thresholds for the
1057 rhesus monkey. *Hearing Research*, **1**.
- 1058 Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K. & Petkov, C.I. (2013)
1059 Auditory artificial grammar learning in macaque and marmoset monkeys. *J Neurosci*, **33**,
1060 18825-18835.

1061

1062

1063