Population Size and Cumulative Cultural Evolution: Fewer Heads

Can Be Better than Many

Nicolas Fay¹, Naomi De Kleine¹, Bradley Walker¹ & Christine A. Caldwell²

¹School of Psychological Science, University of Western Australia, 35 Stirling Highway,

Crawley WA 6009, Australia

²Behaviour and Evolution Research Group, Psychology, School of Natural Sciences,

University of Stirling, Stirling, FK9 4LA, United Kingdom

Running Head: Population Size and Cumulative Culture

Keywords: Demography, Population Size, Cumulative Cultural Evolution, Cultural Evolution,

Innovation, Social Transmission, Success Bias, Transmission Fidelity

Corresponding author:

Nicolas Fay, School of Psychological Science, University of Western Australia

35 Stirling Highway, Crawley, WA 6009 Australia

Email: nicolas.fay@gmail.com; Tel: +61 (0)8 6488 2688; Fax: +61 (0)8 6488 1006

Word count (excluding title page, abstract and references): XXXX words

24 Abstract

25 The extent to which larger populations benefit cumulative cultural evolution (CCE) is
26 contentious. We report a large-scale experiment ($N=543$) that investigates the CCE of
27 technology (paper planes and their flight distance) using a transmission chain design.
28 Population size was manipulated such that participants could learn from the paper planes
29 constructed by 1, 2 or 4 models from the prior generation. These social learning conditions
30 were compared to an asocial Individual Learning condition in which individual participants
31 made repeated attempts at constructing a paper plane, without having access to any planes
32 produced by others. Larger populations generated greater artefact diversity and gave
33 participants access to better-adapted artefacts, but this did not enhance CCE. In fact, there
34 was an inverse relationship between population size and CCE: plane flight distance did not
35 improve over the experimental generations in the 4-Model and 2-Model conditions, but did
36 improve over generations in the 1-Model social learning condition. The incremental
37 improvement in plane flight distance in the 1-Model social learning condition was
38 comparable to that in the Individual Learning condition, highlighting the importance of trial-
39 and-error learning to artefact innovation and adaptation. In the context of this experiment,
40 we conclude that population size alone does not enhance CCE, and that it may be only
41 under certain specific conditions that larger population sizes benefit CCE.

42 Introduction

43 Cumulative cultural evolution (CCE) is the process through which adaptive modifications
44 accumulate over time, driving the incremental improvement of material and symbolic
45 artefacts (Caldwell & Millen, 2009; Fay et al., 2018; Mesoudi & Thornton, 2018; Tomasello,
46 1999). Demography may be crucial to CCE: in particular, larger populations¹ provide access
47 to better-adapted solutions, greater artefact diversity and the opportunity this affords for
48 recombination (Muthukrishna & Henrich, 2016). Henrich (2004) suggests that a decline in
49 the complexity of Tasmanian technology following its isolation from mainland Australia (due
50 to rising sea levels between 12-10 kya) can be explained by the consequent reduction in the
51 population of interacting social learners. However, this explanation is contentious. Vaesen
52 and colleagues (Vaesen, Collard, Cosgrove, & Roebroeks, 2016a) cast doubt on the
53 assumptions behind the mathematical models upon which this explanation is based
54 (Henrich, 2004; see also Powell, Shennan, & Thomas, 2009), and the archeological and
55 ethnographic record it tries to explain (see also Henrich et al., 2016; Vaesen, Collard,
56 Cosgrove, & Roebroeks, 2016b).

57 Laboratory experiments that manipulate population size and test its causal effect on
58 cultural accumulation provide some support for the benefits of larger populations (Derex,
59 Beugin, Godelle, & Raymond, 2013; Kempe & Mesoudi, 2014; Muthukrishna, Shulman,
60 Vasilescu, & Henrich, 2014). For example, having access to more demonstrators helped
61 maintain cultural complexity (image editing and knot tying skill; Muthukrishna et al., 2014),
62 and being a member of a larger group improved artefact performance (on a virtual fishing

¹By which we mean members of an interacting 'effective' population size, as opposed to a
'census' population size in which many of its members are non-interacting.

net and arrowhead production task; Derex et al., 2013). However, other studies find no evidence that a larger population size benefits CCE. For example, the complexity of folk tales was unaffected by population size (Acerbi, Kendal, & Tehrani, 2017), and the CCE of technology (the flight distance of paper planes) was unaffected by the number of artefacts that participants had access to (Caldwell & Millen, 2010). So, the experimental evidence supporting a direct link between population size and CCE is equivocal.

These conflicting experimental results may be explained by the greater opportunity for success-biased copying - the preferential copying of successful models - in some experiments compared to others (see Kendal et al., 2018 for a discussion of the various social learning strategies). In Derex et al. (2013) participants were permitted to view a single artefact produced by a single demonstrator (ranked by their score). In Muthukrishna et al. (2014), participants were informed of the score associated with each artefact, and could view the artefacts produced by any or all of five demonstrators. Participants tended to copy the artefact produced by the most successful demonstrator (Expt. 1), and overwhelmingly chose to view a single artefact produced by a single demonstrator (Expt. 2). So, in both of these studies participants could select the single most successful demonstrator(s) to learn from, and could avoid encountering the artefacts produced by less successful demonstrators. Performance improvements in these experiments may be accounted for by a combination of population size and success-biased copying that was facilitated by the opportunity to filter out less successful variants. So, these experiments do not test the basic effect of population size on CCE.

Success-biased copying may have been more challenging in Caldwell and Millen (2010). Here, participants observed each of the artefacts produced by each of the other members of their group. Participants could therefore not avoid encountering the artefacts

produced by less successful demonstrators via selective filtering. Human working memory - used to hold information for temporary processing - has a limited capacity of around four chunks of information (Cowan, 2001, 2010). In the Caldwell and Millen (2010) experiment, the artefact (paper plane), its performance (flight distance) and its construction method correspond to at least three chunks of information. It follows that as population size was increased (from 1 to 2 to 3 models) participants' working memory capacity may have been overwhelmed, reducing their ability to engage in success-biased copying. In addition, in this experiment, participants were given the same fixed amount of time to study the artefacts produced by the members of their group. This may have disadvantaged participants allocated to the larger populations, as the mean amount of time available to view each artefact decreased as the population size increased.

We present a large-scale experiment ($N= 543$) that tests the effect of population size on the CCE of technology. We used the paradigm developed by Caldwell and Millen (2008, 2009, 2010) because it returns a precise measure of artefact performance (paper plane flight distance), but we adapted the procedure to ensure that the mean viewing time per artefact was the same across the different population sizes. In each population size, participants viewed each of the artefacts produced by each member of the prior experimental generation. So, participants could not avoid encountering less successful artefacts (via filtering), which ensured a 'pure' test of population size. This is crucial to establishing the basic effect of population size on CCE. An asocial, individual learning condition was also included. This condition, often missing from experimental studies of CCE (Miton & Charbonneau, 2018), provided a baseline against which the social learning conditions could be compared. First, we tested if larger populations generated greater artefact diversity (Hypothesis 1) and gave participants access to better-adapted artefacts

(Hypothesis 2), crucial ingredients if larger populations are to benefit CCE. We then tested if larger populations benefited CCE (Hypothesis 3). Finally, we report an exploratory test of the relationship between population size and success-biased copying.

Method

The study received approval from the University of Western Australia Ethics Committee. Participants viewed an information sheet before giving written consent to take part in the study. The information sheet and consent form were both approved by the Ethics Committee. All methods were performed in accordance with the guidelines from the NHMRC/ARC/University Australia's National Statement on Ethical Conduct in Human Research.

Participants

A convenience sample of 543 undergraduate students from the University of Western Australia (391 females) participated in exchange for partial course credit. Participants ranged in age from 19 to 56 years ($M = 24$, $SD = 6.84$).

Task and Procedure

The task for each participant was to construct a paper plane that flew as far as possible using a single piece of A4 paper (Caldwell & Millen, 2008, 2009, 2010). Participants were randomly allocated to an 8-generation transmission chain. This design is widely used to experimentally study social learning (Bebbington, MacLeod, Ellison, & Fay, 2017; Hopper, Flynn, Wood, & Whiten, 2010; Kashima, 2000; Morgan et al., 2015).

Participants were randomly assigned to a condition: Individual Learning, 1-Model, 2-Model or 4-Model. The Individual Learning condition ($N=31$ participants, $N=31$ chains, with the same subject participating eight times, once at each generation) was an asocial learning condition. In the 1-Model ($N=128$ participants, $N=16$ chains, with 1 subject participating once at each generation), 2-Model ($N=128$ participants, $N=8$ chains, with 2 subjects participating once at each generation) and 4-Model ($N=256$ participants, $N=8$ chains, with 4 subjects participating once at each generation) conditions each participant was randomly assigned to a position (1-8) in the chain and completed the task once, within a single generation, before being removed from the experiment. Testing took place in a large room that is 12m long by 7.5m wide. The room allowed us to test up to eight participants simultaneously (separated by visual barriers to eliminate the possibility of social learning occurring from sources other than those intended as part of the experimental design). To enable quick and accurate measurements of flight distance, we placed duct tape across the width of the testing room floor at 1m intervals and marked on each the distance from the participant. Flight distance was measured using a tape measure, from the nearest taped marker to the paper plane.

Each participant received verbal and written instructions prior to commencing the experiment. Participants were seated at a table and given 5 minutes to construct a paper plane that would fly as far as possible using the A4 sheet of paper provided. Plane construction was restricted to folding the paper; ripping the paper or rolling it into a ball was not permitted. A stopwatch was placed on the table so that participants could keep track of the remaining time. In addition, the experimenter verbally informed participants when there was 1 minute remaining, when there were 30 seconds remaining and when their time was up. Next, participants flew their paper plane three times from a seated

position in a high back chair (all chairs were set at an equal height). This was done to reduce the influence of differences in participant height, and to eliminate any 'run-throw' or related techniques, restricting performance, as much as was possible, to the characteristics of the paper plane. For each paper plane, all three flight distances were recorded by the experimenter, and the longest flight distance was recorded alongside the plane on a separate piece of paper. The longest flight distance was representative of the paper plane, correlating strongly with the mean of the other two flight distances, $r(758) = 0.890$, $p < .001$. The experimenter photographed each paper plane, before the plane and its associated maximum flight distance was made available to the next person(s) in the transmission chain.

At Generation 1 all participants constructed a paper plane without any opportunity for social learning. Hence, the conditions were identical at Generation 1. The conditions differed from Generation 2 onwards. In the Individual Learning condition the same participants completed the paper plane task at each experimental generation. At Generation 2, participants in the Individual Learning condition were given 1.5 minutes to study the paper plane they had previously built and its associated maximum flight distance, after which both were removed. Participants were then given 5 minutes to build a plane, and then flew the plane three times. The flight distances were recorded and the maximum flight distance was placed alongside the plane on a separate piece of paper. This procedure was followed eight times, equivalent to the 8 generations in the other conditions. A similar procedure was followed in the 1-Model condition. However, in this condition each participant left the experiment after a single generation testing session and was replaced by a new participant who represented the next generation. The Generation 2 participant studied the paper plane produced by the Generation 1 participant and its associated

maximum flight distance (for 1.5 minutes). They then produced their own paper plane (5 minutes) and flew it three times. The flight distances were recorded by the experimenter, and the plane and its maximum flight distance was made available to the Generation 3 participant and so on across the eight experimental generations.

In the 2-Model condition, Generation 2 participants viewed two planes, each constructed by a different Generation 1 participant, and their associated maximum flight distances. Each plane was separately viewed for 1.5 minutes (i.e., 3 minutes total viewing time) before being removed. Planes were randomly assigned to the participants and rotated every 1.5 minutes. The Generation 2 participants then constructed their own paper plane (5 minutes) and flew it three times. Their planes and their associated maximum flight distances were then passed to the Generation 3 participants. The same procedure was followed by participants in the 4-Model condition, but now with 4 planes and their associated maximum flight distances. Participants in this condition were given 1.5 minutes to separately view each plane (i.e., 6 minutes in total) before constructing a plane (5 minutes) and flying it three times. The planes and their maximum flight distances were then passed to the Generation 3 participants. This procedure continued across the eight experimental generations. The experiment setup is schematically communicated in **Figure 1**.

Figure 1. Experiment setup. Participants were seated at a desk and constructed a paper plane by folding a piece of A4 paper (5 minutes). A stopwatch was placed on the table to indicate the elapsed time. Participants were separated by a visual barrier.

Results

All of the experimental variables were normally distributed (skewness within ± 1.10 and kurtosis within ± 1.60). The data was analysed using linear mixed effects modelling, including by-chain random intercepts. The fixed effects (Condition, Generation) were centered prior to analysis². All analyses were performed and all figures were created in R (R Core Team, 2013). Statistical models were estimated using the `lmer()` function of `lme4` (Bates, Maechler, Bolker, & Walker, 2013).

Section 1 tests that the prerequisites for larger populations to benefit cumulative

²The Individual Learning, 1-Model, 2-Model and 4-Model conditions were numerically coded: 0, 1, 2 and 4 respectively, the numeric value corresponding to the number of other models the participant could socially learn from.

cultural evolution (CCE) were met (i.e., access to increased artefact diversity and better-adapted artefacts). Section 2 tests if larger populations benefited CCE. Section 3 reports an exploratory test of the relationship between population size and success-biased copying. Specifically, it tests the extent to which the opportunity for success-biased copying in larger populations was counteracted by constraints on human working memory.

Section 1. Prerequisites for Larger Populations to Benefit CCE

Larger populations can benefit CCE if they provide access to greater artefact diversity (Hypothesis 1), and better-adapted artefacts (Hypothesis 2). Artefact diversity was quantified by determining the range of flight distances in the 2-Model and 4-Model conditions (maximum recorded flight distance minus minimum recorded flight distance) that each participant had access to at each generation. Because participants in the Individual Learning and 1-Model conditions were exposed to a single plane, there was no diversity in plane performance. Artefact quality was quantified by determining the maximum flight distance of the paper planes participants constructed at Generation 1 in each condition. In the Individual Learning and 1-Model conditions this was the plane built by each participant at Generation 1. In the 2- and 4-Model conditions this was the furthest-flying plane of the 2 or 4 planes produced by participants at Generation 1. Examining plane flight distance at Generation 1 ensured the artefacts were unaffected by the social learning condition.

As predicted, artefact diversity was higher in the 4-Model condition compared to the 2-Model condition. There was no evidence of a statistical change in artefact diversity over the experimental generations (see **Table 1**, **Figure 2A**). As predicted, the maximum flight distance increased as population size increased ($\beta = 0.579$, $SE = 0.156$, $t = 3.713$, $p < .001$)

(see **Figure 2B**). The prerequisites for larger populations to benefit CCE were met: larger populations gave participants access to more artefact diversity, supporting Hypothesis 1, and better-adapted artefacts (at Generation 1), supporting Hypothesis 2.

Table 1. Diversity in plane flight distance over generations in the 2- and 4-Model conditions: Results from the linear mixed effects model.

Fixed Effects	Estimate	Standard Error	t value	Pr(> t)
Initial linear mixed effects model (BIC = 551.3)				
Condition	0.854	0.179	4.768	<.001
Generation	0.039	0.071	0.553	.582
Condition*Generation	-0.056	0.071	-0.796	.428
Final linear mixed effects model (BIC = 542.5)				
Condition	0.854	0.179	4.768	<.001

Figure 2. Panel A shows the diversity in plane flight distances (maximum – minimum) across the experimental generations in the 2- and 4-Model conditions (plotted for each chain). The dot points reflect the overall mean at each generation. The red straight line is the linear

model fit and the light red shaded area is the 95% confidence interval. Panel B shows the maximum plane flight distance at Generation 1 in the Individual Learning, 1-, 2- and 4-Model conditions. The dot points reflect the mean for each chain. Error bars are the 95% confidence intervals.

Section 2. Do Larger Populations Benefit CCE?

Twenty-five outliers (3.29% of data) were identified using the Interquartile Range rule (see Moore, McCabe, & Craig, 1993). These values were replaced by the value corresponding to the 95th percentile, done at each level of generation. The best fitting model specified Condition and Generation as fixed effects with interaction (see **Table 2**, **Figure 3A**). Comparison of the Individual Learning and 1-Model conditions indicated an increase in flight distance over the experimental generations (by 1.38m and 1.22m from Generation 1 to 8 respectively; $\beta = 0.172$, $SE = 0.035$, $t = 4.796$, $p < .001$; see **Figure 3B**). By contrast, there was no statistical evidence of an increase in flight distance in the 2- and 4-Model conditions (0.22m and -0.23m from Generation 1 to 8 respectively; $\beta = -0.020$, $SE = 0.045$, $t = -0.456$, $p = .649$; see **Figure 3B**). Larger populations gave participants access to greater diversity in paper plane performance and better-adapted planes (at Generation 1), but did not enhance CCE. In fact, larger populations negatively affected CCE. We therefore reject Hypothesis 3.

Table 2. Change in flight distance over generations: Results from the linear mixed effects model.

Fixed Effects	Estimate	Standard Error	t value	Pr(> t)
---------------	----------	----------------	---------	----------

Condition	-0.003	0.099	-0.034	.973
Generation	0.075	0.029	2.574	.010
Condition*Generation	-0.063	0.017	-3.657	<.001

Figure 3. Panel A shows the change in plane flight distances across the experimental generations in the different conditions (plotted for each chain). The dot points reflect the overall mean at each generation. The red straight line is the linear model fit and the light red shaded area is the 95% confidence interval. Panel B shows the overall change in plane flight distance across the experimental generations (Generation 8 – Generation 1) for each condition. The dot points reflect the mean for each chain. Error bars are the 95% confidence intervals.

Section 3. Why Did Larger Population Size Not Benefit CCE?

Success-biased copying predicts that the degree to which a variant is copied will be associated with its quality. However, if working memory is overwhelmed by the greater number of variants participants have access to in larger populations, this will lower transmission fidelity and weaken success-biased copying. This hypothesis was tested by

using plane flight distance (at Generation N) to predict plane similarity (at Generation N and Generation N+1).

Photographs of each plane were taken from the same angle at each generation. Pairs of photographs at Generation N and Generation N+1 were presented on a computer screen and rated (by author NDK) for similarity on a 10 point Likert-type scale, from 0 = *extremely dissimilar* to 9 = *practically identical*. In total, 1449 pairs of photographs were rated for similarity. A subset (678 pairs of planes; randomly sampled from each condition) was also rated for similarity by a second judge. The raters were blind to the condition or the generation the planes were sampled from. Comparison of the two sets of plane similarity ratings showed acceptable inter-coder agreement, $r(677) = 0.588$, $p < .001$.

The plane similarity ratings were entered into a linear mixed effects model with Condition and Generation N plane flight distance entered as fixed effects. The best fitting model specified Condition and Generation N plane flight distance as fixed effects without interaction (see **Table 3**, **Figure 4**). Consistent with success-biased copying, plane flight distance (at Generation N) predicted plane similarity (Generation N and Generation N+1). The absence of an interaction effect indicates that the influence of success-biased copying was comparable across the experimental conditions. Increased population size lowered transmission fidelity; as population size increased plane similarity between the adjacent generations decreased (see **Figure 4B**). This supports our exploratory hypothesis that working memory constraints weakened copying fidelity as population size increased.

Table 3. The relationship between plane flight distance (at Generation N) and plane similarity (at Generation N and Generation N+1) across the experimental conditions: Results from the linear mixed effects model.

Fixed Effects		Estimate	Standard Error	t value	Pr(> t)
Initial linear mixed effects model (BIC = 6436.9)					
Condition		-0.318	0.058	-5.457	<.001
Generation N Distance		0.247	0.030	8.286	<.001
Condition*Generation	N	-0.028	0.020	-1.401	.162
Distance					
Final linear mixed effects model (BIC = 6431.6)					
Condition		-0.317	0.059	-5.374	<.001
Generation N Distance		0.242	0.030	8.177	<.001

Figure 4. Panel A shows the relationship between plane flight distance (in metres, at Generation N) and plane similarity (Generation N and Generation N+1) in the different conditions. The red straight line is the linear model fit and the light red shaded area is the 95% confidence interval. Panel B shows the mean plane similarity ratings in the different conditions (plotted for each participant). Error bars are the 95% confidence intervals.

318 Discussion

319 The influence of population size on cumulative cultural evolution (CCE) is contentious
320 (Henrich et al., 2016; Vaesen et al., 2016b, 2016a). Using a large-scale experiment, in which
321 participants were unable to filter out less successful artefacts, we investigated the ‘pure’
322 effect of population size on the CCE of technology (paper planes). As predicted, larger
323 populations generated greater artefact diversity – conferring more opportunity for artefact
324 recombination – and access to better-adapted artefacts. However, this did not benefit CCE.
325 In fact, there was an inverse relationship between population size and CCE, such that CCE
326 was not observed in the 4- and 2-Model social learning conditions, but was observed in the
327 1-Model social learning condition. Performance improvements in the 1-Model social
328 learning condition were comparable to those in the Individual Learning condition. We then
329 tested the speculative hypothesis that working memory constraints may have reduced
330 copying fidelity, weakening success-biased copying in the larger populations. This
331 hypothesis was supported; success-biased copying was evident in all conditions, but overall
332 copying fidelity decreased as population size increased. Equally, lower copying fidelity in
333 the multiple models conditions may be due to artefact recombination. Further research is
334 needed to tease apart these competing explanations.

335 The finding that increasing population size did not improve CCE is consistent with
336 research on group performance and decision-making (Kerr & Tindale, 2004). As group size is
337 increased, so too is the coordination problem faced by its members, which can lower group
338 productivity relative to its potential productivity (Hill, 1982; Ingham, Levinger, Graves, &
339 Peckham, 1974; Steiner, 1972). This is clearly seen in brainstorming groups, where the need
340 to coordinate idea production - so as not to speak over the top of others - causes the mean

number of ideas generated per person to decrease as group size is increased (Gallupe et al., 1992). The absence of a coordination problem in the Individual Learning condition may help explain why participants allocated to this condition performed as well as those allocated to the 1-Model social learning condition, and better than those allocated to the 2- and 4-Model social learning conditions. As advised by Miton & Charbonneau (2018), the Individual Learning condition and the social learning conditions were matched in terms of total learning time. However, the conditions are mismatched in terms of the participants' task experience; participants in the Individual learning condition constructed a paper plane eight times, whereas participants in the social learning conditions constructed a paper plane once over the course of the experiment. We therefore recommend caution about direct comparisons between the Individual Learning condition and the social learning conditions. Our results nevertheless emphasize the importance of individual trial-and-error learning to artefact innovation and adaptation.

The findings of the present experiment do not preclude the importance of population size to CCE. However, in the context of the experimental paradigm used – a design that included relatively brief and sequential inspection of the artefacts produced by members of the prior generation – our findings indicate that a larger population size alone may not be sufficient to benefit CCE. In fact, our results suggest that larger populations can be deleterious to CCE due to the working memory constraints and coordination problems inherent to larger groups. We propose that it may be only under certain specific conditions that larger population sizes benefit CCE. Network structure, or connectivity, is likely to be crucial (Centola & Baronchelli, 2015; Derex & Boyd, 2015, 2016; Fay, Garrod, Roberts, & Swoboda, 2010). Although people regularly interact in groups, the basic arena for social interaction is the dyad (Clark, 1996). A major advantage of dyadic interaction is that it

minimizes the aforementioned working memory and coordination constraints, allowing participants to dedicate their cognitive resources to evaluating the information presented. Dyadic interaction may be the dominant context for social interaction because it is optimally adapted for information exchange.

Language evolution experiments suggest that larger population sizes benefit CCE when member interactions are organized in dyads. The communication systems that evolve in larger populations (organized as a series of dyadic interactions) are optimized for effective and efficient communication, and are easier to acquire and are transmitted with higher fidelity by subsequent generations compared to the communication systems that evolve in smaller populations (Fay & Ellison, 2013; Fay, Garrod, & Roberts, 2008; Fay et al., 2010). In these experiments, because participants were interacting in dyads, they had the cognitive resources available to assess the informational value of the communication variants they encountered, and adopt or discard them based upon this evaluation. Iterated across dyadic interactions in a large population, this copy-if-better strategy (Rogers & Fay, 2016; Schlag, 1998, 1999) led to the survival and propagation of communication variants that were optimally adapted for communication use and acquisition (Tamariz, Ellison, Barr, & Fay, 2014). A similar benefit of larger populations is observed in a naturalistic study of linguistic systems (Lupyan & Dale, 2010).

Conclusion

We conducted a large-scale experiment ($N=543$) that tested the effect of population size on the cumulative cultural evolution (CCE) of technology (paper planes). Larger populations gave participants access to greater diversity in paper plane performance and better-adapted

planes, but it did not enhance CCE. In fact, there was an inverse relationship between population size and CCE: there was no improvement in plane flight distance over the experimental generations in the 4-Model and 2-Model social learning conditions, but there was an improvement in the 1-Model social learning condition. The improvement in plane flight distance in the 1-Model social learning condition was comparable to that in the asocial Individual Learning condition, highlighting the importance of individual trial-and-error learning to artefact innovation and adaptation. An exploratory analysis indicated that having access to more artefacts may have overwhelmed participants' working memory in the larger populations, reducing their ability to learn from the more successful artefacts. In the context of the present experiment, a larger population size did not benefit CCE, indicating that other factors may also be important for population size to benefit CCE.

References

- Acerbi, A., Kendal, J., & Tehrani, J. J. (2017). Cultural complexity and demography: The case of folktales. *Evolution and Human Behavior*, 38(4), 474–480. <https://doi.org/10.1016/j.evolhumbehav.2017.03.005>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, 1(4).
- Bebbington, K., MacLeod, C., Ellison, T. M., & Fay, N. (2017). The sky is falling: evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior*, 38(1), 92–101. <https://doi.org/10.1016/j.evolhumbehav.2016.07.004>
- Caldwell, C. A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, 165–171.

- 411 Caldwell, C. A., & Millen, A. E. (2009). Social Learning Mechanisms and Cumulative Cultural
412 Evolution: Is Imitation Necessary? *Psychological Science*, 20(12), 1478–1483.
- 413 Caldwell, C. A., & Millen, A. E. (2010). Human cumulative culture in the laboratory: Effects of
414 (micro) population size. *Learning & Behavior*, 38(3), 310–318.
- 415 Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An
416 experimental study of cultural evolution. *Proceedings of the National Academy of*
417 *Sciences*. <https://doi.org/10.1073/pnas.1418838112>
- 418 Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- 419 Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of
420 mental storage capacity. *Behavioral and Brain Sciences*, 24(01), 87–114.
- 421 Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and
422 why? *Current Directions in Psychological Science*, 19(1), 51–57.
- 423 Derex, M., Beugin, M.-P., Godelle, B., & Raymond, M. (2013). Experimental evidence for the
424 influence of group size on cultural complexity. *Nature*, 503(7476), 389–391.
425 <https://doi.org/10.1038/nature12774>
- 426 Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature*
427 *Communications*, 6, 8398. <https://doi.org/10.1038/ncomms9398>
- 428 Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within
429 groups. *Proceedings of the National Academy of Sciences*, 113(11), 2982–2987.
- 430 Fay, N., & Ellison, T. M. (2013). The Cultural Evolution of Human Communication Systems in
431 Different Sized Populations: Usability Trumps Learnability. *PLoS ONE*, 8(8), e71781.
432 <https://doi.org/10.1371/journal.pone.0071781>
- 433 Fay, N., Ellison, T. M., Tylén, K., Fusaroli, R., Walker, B., & Garrod, S. (2018). Applying the
434 cultural ratchet to a social artefact: The cumulative cultural evolution of a language

- 435 game. *Evolution and Human Behavior*, 39(3), 300–309.
 436 <https://doi.org/10.1016/j.evolhumbehav.2018.02.002>
- 437 Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved
 438 communication systems. *Philosophical Transactions of the Royal Society B-Biological*
 439 *Sciences*, 363(1509), 3553–3561.
- 440 Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human
 441 communication systems. *Cognitive Science*, 34(3), 351–386.
- 442 Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker, J.
 443 F. (1992). Electronic Brainstorming And Group Size. *Academy of Management*
 444 *Journal*, 35(2), 350–369. <https://doi.org/10.5465/256377>
- 445 Henrich, J. (2004). Demography and Cultural Evolution: How Adaptive Cultural Processes can
 446 Produce Maladaptive Losses: The Tasmanian Case. *American Antiquity*, 69(2), 197–
 447 214. <https://doi.org/10.2307/4128416>
- 448 Henrich, J., Boyd, R., Derex, M., Kline, M. A., Mesoudi, A., Muthukrishna, M., ... Thomas, M.
 449 G. (2016). Understanding cumulative cultural evolution. *Proceedings of the National*
 450 *Academy of Sciences*, 113(44), E6724–E6725.
 451 <https://doi.org/10.1073/pnas.1610005113>
- 452 Hill, G. W. (1982). Group versus individual performance: Are N+ 1 heads better than one?
 453 *Psychological Bulletin*, 91(3), 517–539.
- 454 Hopper, L. M., Flynn, E. G., Wood, L. A., & Whiten, A. (2010). Observational learning of tool
 455 use in children: Investigating cultural spread through diffusion chains and learning
 456 mechanisms through ghost displays. *Journal of Experimental Child Psychology*,
 457 106(1), 82–97.

- 458 Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies
459 of group size and group performance. *Journal of Experimental Social Psychology*,
460 10(4), 371–384. [https://doi.org/10.1016/0022-1031\(74\)90033-X](https://doi.org/10.1016/0022-1031(74)90033-X)
- 461 Kashima, Y. (2000). Maintaining cultural stereotypes in the serial reproduction of narratives.
462 *Personality and Social Psychology Bulletin*, 26(5), 594–604.
- 463 Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the effect of group
464 size on cultural accumulation. *Evolution and Human Behavior*, 35(4), 285–290.
465 <https://doi.org/10.1016/j.evolhumbehav.2014.02.009>
- 466 Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018).
467 Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive*
468 *Sciences*, 0(0). <https://doi.org/10.1016/j.tics.2018.04.003>
- 469 Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review*
470 *of Psychology*, 55, 623–655.
- 471 Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure.
472 *PLoS ONE*, 5(1), e8559. <https://doi.org/10.1371/journal.pone.0008559>
- 473 Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proc. R. Soc. B*,
474 285(1880), 20180712. <https://doi.org/10.1098/rspb.2018.0712>
- 475 Miton, H., & Charbonneau, M. (2018). Cumulative culture in the laboratory: methodological
476 and theoretical challenges. *Proc. R. Soc. B*, 285(1879), 20180677.
477 <https://doi.org/10.1098/rspb.2018.0677>
- 478 Moore, D. S., McCabe, G. P., & Craig, B. A. (1993). *Introduction to the Practice of Statistics*.
479 New York: W. H. Freeman.
- 480 Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M.,
481 ... others. (2015). Experimental evidence for the co-evolution of hominin tool-making

- 482 teaching and language. *Nature Communications*, 6(6029).
483 <https://doi.org/10.1038/ncomms7029>
- 484 Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Phil. Trans. R. Soc.*
485 *B*, 371(1690), 20150192.
- 486 Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2014). Sociality influences
487 cultural complexity. *Proceedings of the Royal Society of London B: Biological*
488 *Sciences*, 281(1774), 20132511.
- 489 Powell, A., Shennan, S., & Thomas, M. G. (2009). Late Pleistocene demography and the
490 appearance of modern human behavior. *Science*, 324(5932), 1298–1301.
- 491 R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna,
492 Austria: R Foundation for Statistical Computing. Retrieved from [http://www.R-](http://www.R-project.org/)
493 [project.org/](http://www.R-project.org/)
- 494 Rogers, S. L., & Fay, N. (2016). Stick or Switch: A Selection Heuristic Predicts when People
495 Take the Perspective of Others or Communicate Egocentrically. *PloS One*, 11(7),
496 e0159570.
- 497 Schlag, K. H. (1998). Why imitate, and if so, how?: A boundedly rational approach to multi-
498 armed bandits. *Journal of Economic Theory*, 78(1), 130–156.
- 499 Schlag, K. H. (1999). Which one should I imitate? *Journal of Mathematical Economics*, 31(4),
500 493–522.
- 501 Steiner, I. (1972). *Group process and productivity*. New York: Academic Press.
- 502 Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the
503 evolution of human communication systems. *Proceedings of the Royal Society B:*
504 *Biological Sciences*, 281(1788), 1–6. <https://doi.org/10.1098/rspb.2014.0488>

- 505 Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard
506 University Press.
- 507 Vaesen, K., Collard, M., Cosgrove, R., & Roebroeks, W. (2016a). Population size does not
508 explain past changes in cultural complexity. *Proceedings of the National Academy of*
509 *Sciences of the United States of America*, 113(16), E2241-2247.
510 <https://doi.org/10.1073/pnas.1520288113>
- 511 Vaesen, K., Collard, M., Cosgrove, R., & Roebroeks, W. (2016b). Reply to Henrich et al.: The
512 Tasmanian effect and other red herrings. *Proceedings of the National Academy of*
513 *Sciences*, 113(44), E6726–E6727. <https://doi.org/10.1073/pnas.1613074113>
514