The “Color Game” App

General presentation paper

This presentation paper for the Color Game App accompanies six pre-registered study plans that have been registered, along with this paper, on the 11th of April, 2018.

Olivier Morin*, James Winters*, Thomas Müller*, Tiffany Morisseau°, Christian Etter ˙, Simon Greenhill*.

* Max Planck Institute for the Science of Human History, Department of Linguistic and Cultural Evolution, Jena, Germany.
° Central European University, Cognitive Science Department, Budapest, Hungary.
˙ Etter Studios, Zurich, Switzerland.

Summary. Unlike a standard online experiment, a gaming app lets participants interact freely with a vast number of partners, as many times as they wish. The gain is not merely one of statistical power. Cultural evolutionists can use gaming apps to allow large numbers of participants to communicate synchronously; to build realistic transmission chains that avoid the losses of information that occurs in linear chains; to study the effects of partner choice as well as partner control in social interactions. We illustrate these methodological opportunities by presenting the “Color Game” app. Built around a referential communication game where players must communicate a target colour using black and white symbols, the game allows large numbers of players to interact freely and build shared visual languages. By assigning players randomly to evolving sub-populations, the app can simulate the demographic dynamics underlying language divergences, providing an experimental test for linguistic and cultural phylogenies.

A. Three problems for language evolution that only an app can solve

Calls for a "smartphone psychology" (Miller 2012) or a "computational cognitive revolution" (Griffiths 2015) regularly invite cognitive and social scientists to make use of the new tools offered by so-called "pervasive ubiquitous computing". Apps and social networks, in contrast to more standard digital tools like online questionnaires or libraries, allow researchers to collect abundant data on a long-term basis, getting access to participant behaviour almost around the clock. This has made them popular in psychology, chiefly among clinicians who can use them to help and monitor patients in ways that would be too costly for standard human interventions (Fitzpatrick, Darcy, and Vierhile 2017). In comparison with this therapeutic use, smartphone apps designed to generate data for fundamental research are rather under-developed and under-used in the behavioural sciences. Promising exceptions have been developed for cognitive psychology, for psychometric purposes or to document demographic profiles for various standard psychological experiments (Lathia et al. 2017; Brown et al. 2014; Klindt, Devaine, and Daunizeau 2017).

Why have app-based experiments not been more widely adopted? The data generated by apps and social networks are usually proprietary and are increasingly less accessible to research as users demand more privacy and companies start monetising their data. The growing displacement of text by videos and pictures also means the data generated by apps and social networks are becoming “unstructured” (*sensu* Gandomi & Haider, 2015): they increasingly require special mining techniques like automated image or speech recognition to become analysable. Faced with these issues, most researchers have turned to online experiments or questionnaires, often relying on cheap labour from crowdsourcing services such as Amazon's “mechanical turk”. Doing so allows scientists to scale up their experiment and gain statistical power, without damage to the quality of the resulting data (Casler, Bickel, and Hackett 2013). Satisfactory though it may be for most researchers, the crowdsourcing of experiments arguably misses quite a few opportunities afforded by the digital revolution.

Experimental language evolution is a case in point. It aims to reproduce some features of language evolution, under laboratory condition and using artificial languages, usually generated by participants playing referential communication games where often the synthetic languages go through transmission chains (Scott-Phillips and Kirby 2010). This method yields robust, replicable claims that naturalistic observations would reach with difficulties (Kirby et al. 2015). However, arbitrary constraints limit the method's capacity to simulate linguistic interactions in a realistic way, even with crowdsourcing. App-based data collection may lift three such limitations.

- *The trade-off between synchrony and large samples.* Synchronous interaction is a crucial dimension of most human communication (Enfield 2017). Key features of linguistic interactions break down when two interlocutors cannot be present inside the same timeframe, like turn-taking and repair (Clark and Wilkes-Gibbs 1986; Dingemanse et al. 2015). Most language evolution experiments can only study synchronous interactions on a very small scale due to practical restrictions on the number of participants that can be studied. An app-based approach can lift this burden.

- *Closed vs. open transmission chains*. Transmission chains are a key aspect of experimental language evolution and cultural evolution research since the 1930s (Bartlett 1932; Mesoudi and Whiten 2008). The standard and widely emulated transmission chain set-up is a game of Telephone where a participant A is asked to transmit a content to B, then B is asked to do the same with C, and so on. Each subject is typically required to transmit a cultural content to one other subject, contrary to real life where agents freely choose whether to transmit a given content, and how many people to transmit it to (Morin, 2015: 122–130). Linear transmission chains, where each participant has exactly one interaction with one other participant, invariably lead to severe losses as information from upstream chain cannot be recovered (Claidière & Sperber, 2010, Morin 2015). Lastly, the generational turn-over simulated by laboratory chains is a far cry from realistic demographic dynamics. Solving these issues require open-ended transmission chains, but these are difficult to create and monitor in controlled conditions outside the lab, where strict limits are imposed on the number of participants.

- *Partner control vs. partner choice*. Language evolution experiments invariably pair participants with one or a few partners they must communicate with. A participant faced with an obtuse partner has no choice but to try and teach them better tools for communication; finding a more suitable partner is not an option. A parallel with the “evolution of cooperation” literature seems apt here. Experimental studies of helping, cooperation, or reciprocity among humans are dominated by economic games where two randomly assigned partners must choose how to share a resource or whether to cooperate to generate a common good. Success or failure depends entirely on "partner control", behavioural ecologists' name for all the actions that encourage or discourage a given partner from cooperating (Noë and Hammerstein 1994). Yet real-life cooperation also hinges on "partner choice", the switch from less cooperative partners to more cooperative ones. A "market for cooperation" thus allows the most reliable partners to co-opt one another. Taking this dimension into account has recently lead to important theoretical advances in the study of human cooperation (Baumard, André, and Sperber 2013). Introducing partner choice in experimental language evolution may lead to similarly momentous changes, but here again the practical constraints of experiments (crowdsourced or lab-based) stand in the way.

B. The “Color Game” App

The Color Game app is a tool designed to resolve these issues by experimenting on language at a large scale, while allowing large numbers of participants to interact synchronously as often as they choose and with a wide variety of partners. At its core is a standard referential communication task (Yule 2013). One player (the Sender) is presented with a "target colour", indicated by a dot. This colour corresponds to one of the colours in an array of four that is seen by the other player (the Receiver). The sender's goal is to communicate with the Receiver to help them pick the target colour, earning points. Senders must communicate using black and white symbols that bear no straightforward association with any single hue of colour (Figs. 1 & 2). These symbols have been experimentally tested to make sure that they would be neither too easy (evoking too narrow a range of colours), nor too difficult (allowing no colour associations whatsoever). Laboratory experiments show that the symbols are as ambiguous as desired, since different pairs of participants can use them to solve the communication task above chance, but distinct pairs will associate the same symbol with different colours (Müller et al. in prep.). Independently of any communication task (in the lab or with the app), we have collected data on prior associations between the game's 35 symbols and its 32 colours, by asking 647 participants to freely associate one given symbol with a colour, or the reverse, in one-shot tasks. When confronted with the data generated by the app, this prior association data will allow us to observe and quantify the extent to which communicative conventions may strengthen or override a symbol's pre-existing meanings. To maximise the variability in symbol use, as well as provide the game with a reward structure, the players who start the game are only provided with a random sample of 10 symbols (out of 35), earning the right to use additional symbols progressively as they earn points and ascend to new levels.

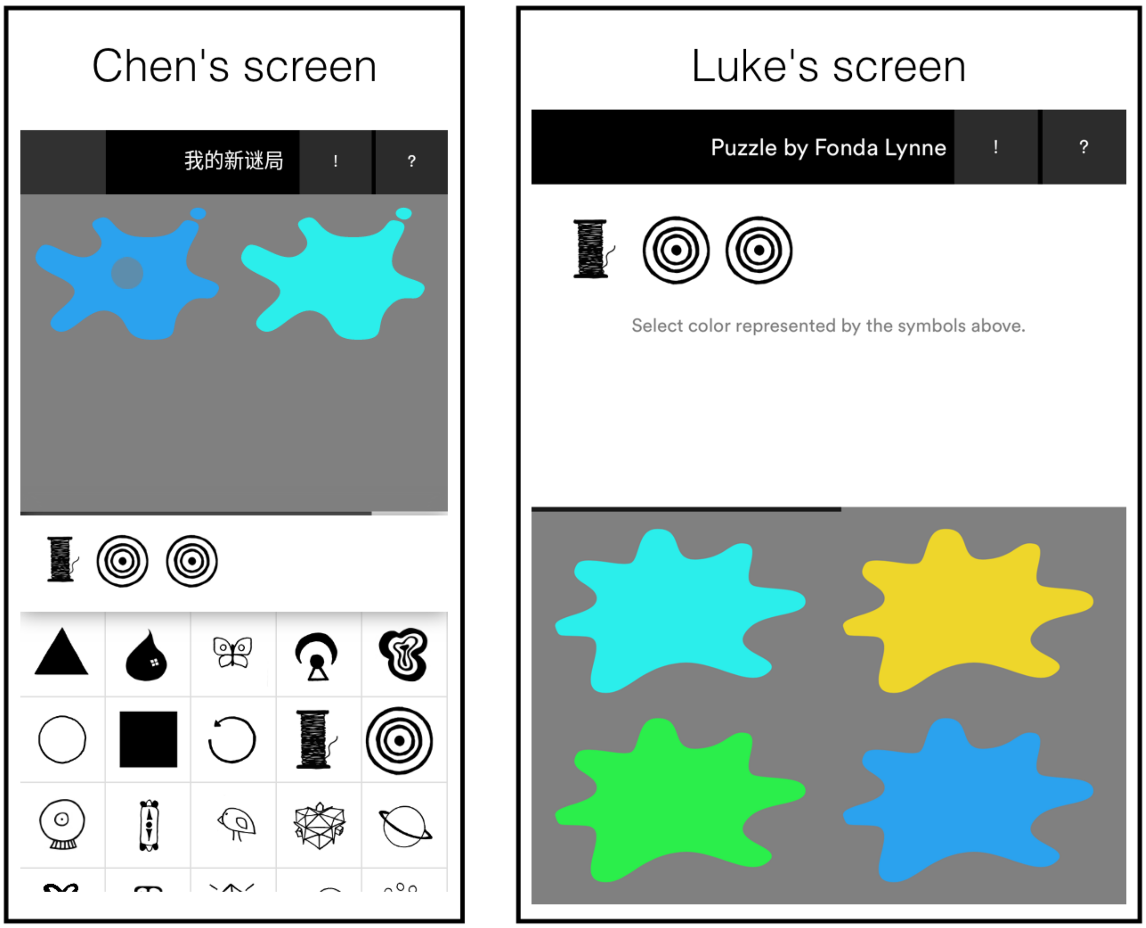

Fig. 1. A trial in synchronous mode. Chen (Sender) communicates with Luke (Receiver) to help Luke find the target colour (here, the darker shade of blue), marked for Chen by a dot. (Screen capture from a Beta test, using tutorial symbols instead of the normal symbols.)

Unlike most language evolution experiments, our app does not provide players with trial-by-trial feedback on the success or failure of communication. A block of 10 trials must be played by both Sender and Receiver for either of them to earn their points. After every block, the Receiver is told how many of the 10 they got right, but not which ones. If both players were playing synchronously (see below) both players get that information, otherwise the Sender merely knows that someone played with their messages. Our reason to avoid trial-by-trial feedback is that it would let Receivers know instantly which symbol their Sender associates with which colour, allowing Receivers to learn a Sender's code by mere association. Instead, players must leverage the symbols' pre-existing connotations (vague as they are) to build shared conventions. Our laboratory experiments (Mueller, Winters, and Morin, n.d.) show that most participant pairs play above chance, and that all above-chance participants achieve significant progress with time (implying that the symbols acquire informative meanings they lacked at the start of the game).

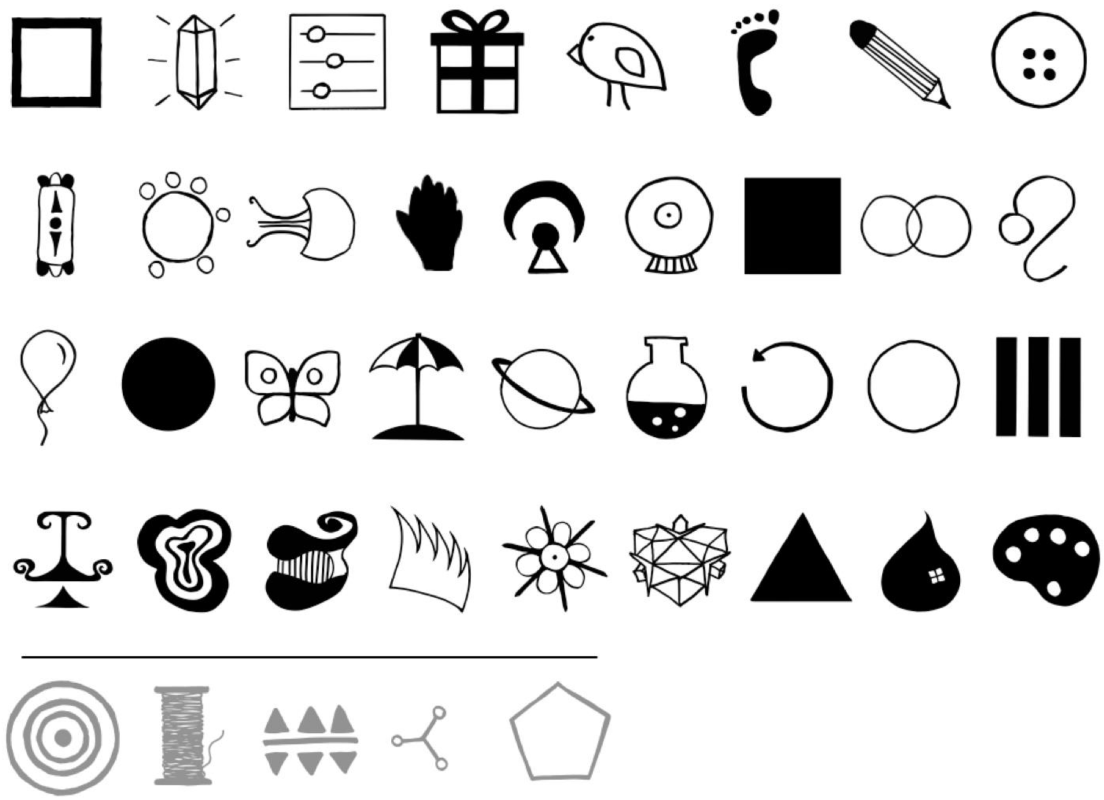

Fig. 2. The 35 symbols used in the game (first four rows). Bottom row, in grey: the five symbols used for the tutorial and for the videos advertising the game (these symbols are for tutorials only).

The game's colour space is designed to make all trials as comparable as possible, save for one randomized intervention. Each of the game's 32 colours is drawn from the CIE2000 colour space (Luo, Cui, and Rigg 2001), chosen because it provides a metric for distance between colour hues ("Delta E") that was built to reflect perceptual distance, as opposed to merely physical quantities. The colours are equal in luminance (L = 55) and saturation (S = 85), with a constant perceptual distance between any colour and its two neighbours of Delta E = 7.8 (Fig. 3). Thirty-two colour arrays are formed from this set of 32 colours by picking every fourth colour along the dimension of hue, until a four-colours array is formed, using each of the 32 colours as starting point (Fig. 4.). This way, all colours occur in exactly four arrays. The arrays are randomly generated in terms of what portion of a Receiver's array is visible to a Sender. In addition to the target colour, a Sender may see some or all of the colours visible in Receiver's array. This quantity varies from one (only the target) to four (the full array).

Fig. 3. The game's colour space. Each colour is given its associated Hex code (as used by the app).

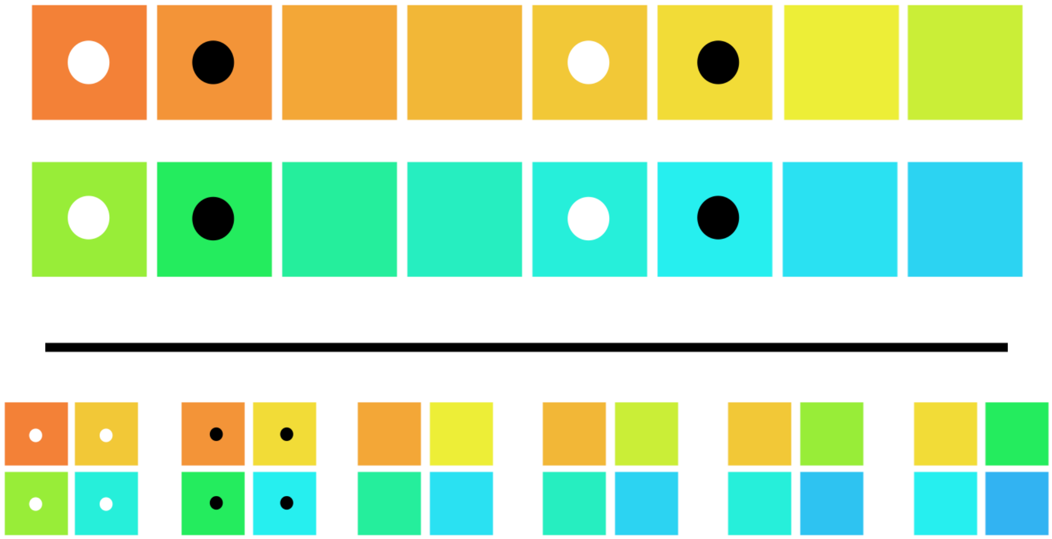

Fig 4. How colour arrays are built. Top row: The composition of two colour arrays, one marked by white dots, the other by black dots, is shown relative to the colour space. Bottom row: Six contiguous colour arrays (out of 32), including the white-dot and black-dot ones.

The app leaves players free to choose their interlocutors and the format of their interaction. It allows for both “synchronous” (i.e., live, or in real time) and “asynchronous” play. Senders playing asynchronously simply type black and white symbols corresponding to the target colour. The symbols are then sent to Receivers along with the colour array, and Receivers figure out the target from these symbols. The message will always be accompanied by the corresponding array of four colours, out of which the Receiver must pick the target. These *asynchronous puzzles* remain available inside the app indefinitely, but each puzzle disappears as soon as one Receiver has played it.

Synchronous play requires the two players to contact each other and stay connected for as long as they play; this enables them to communicate in real time, and exchange repair signals consisting of the signs "?" and "!". The players are not told what these punctuation marks precisely mean in the context of the game, and we expect variation in the way they will be used. Asynchronous play makes it possible for players to interact with a vast number of players at the time of their choosing, greatly enhancing the number of interactions we can observe, while synchronous play allows us to study interactional properties of communication, such as repair.

The app's players are also free to choose their partners from a vast pool of players (Fig. 5). To play with another contact, a player either invites them for synchronous play and waits for the invitation to be accepted, or sends an asynchronous message, which may be broadcast to the whole group or sent to a specific individual contact who may open it at a later time. To ensure that the app always contains a sufficient number of high-quality asynchronous puzzles, Receivers must pay a number of points to the Sender whose puzzle they want to play. The Sender receives these points regardless of the Receiver's performance, incentivising participants to play as Senders. Extra points may be earned by playing an especially difficult, time-limited "speed mode", available to players after a certain level.

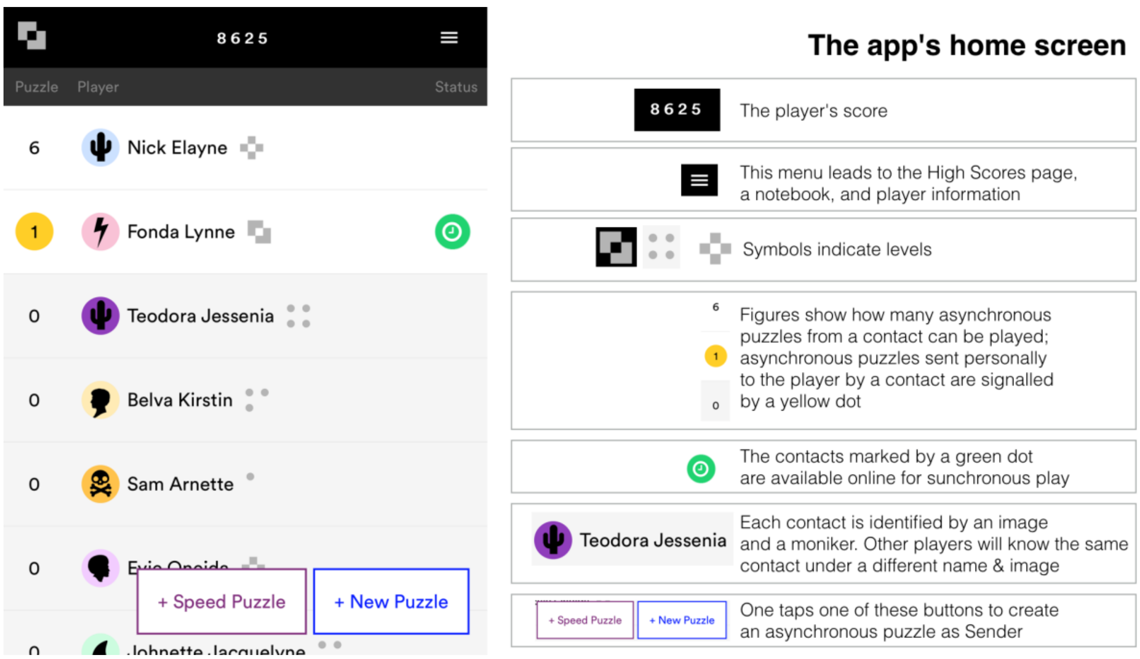

Fig. 5. The app's home screen (left,) with legend (right). The colourful logos that identify each contact are randomly generated from a set of black and white pictures and a set of colours.

We expect to see the emergence of a market dynamic, whereby the most skilful players get more proposals for synchronous play and more takers for their asynchronous puzzles. The app makes this more likely by putting each player's last-played-with contacts on top of the list, showing their level, and attaching a distinctive picture to each. The pricing of puzzles, coupled with quality indicators, is meant to discourage the creation of low-quality puzzles or the absorption of high-quality puzzles by negligent players.

Strict player anonymity is ensured by a system of pseudonyms and "cover names". The app records no personal identifiers such as names, but gives each player a public pseudonym, displayed on the "top players" score board. Crucially, however, that pseudonym cannot be used to identify a player as a contact to play with. Instead, contacts are only ever known to one another by means of "cover names" that vary from contact to contact and from player to player. Player X knows her contact through a list of randomly generated names that reliably identify contact A, contact B, contact C, and so on. Player Y can also identify the same contacts A, B, C, but by completely different names, so that A's name for X does not correspond in any way to A's name for Y, even though both players can reliably identify and contact A. This ensures that players cannot circumvent the app to contact one another in real life, on social networks, etc.

The Color Game allows us to divide our pool of players into an indefinite number of sub-pools. Players are divided, when the app is launched, between two big pools of players, with each new player being assigned randomly to one half. All interactions between players are restricted to their half, and this primary division stays in place indefinitely. We will further divide our pool of players into more plastic sub-groupings, to create population divergences or merging events. The app lets us create such sub-groupings quite freely, and merging two subgroups back into one group is just as easy. A player assigned to a new subgroup loses contact with her out-group contacts but retains her in-group ones. New players are randomly assigned to one of the new subgroups. When two sub-groups are merged, all their players become free to play with one another. This grouping functionality allows us to track groups over time, and modify their evolutionary trajectories, mimicking the evolution of languages (see section D).

Participants agree to have their data collected in anonymous format and for research purposes in a consent form approved at the start of the game. The form and the app itself were approved by the Max Planck Society headquarters' ethical committee (advice n° 2017_05). The app's source code is open and the anonymous data will be made available after a period of embargo. In addition to recording all the players' moves as Sender or Receiver, the app also records notes taken by the players who wish to make use of "the notebook", a feature of the app that allows players to write down the meaning of a particular symbol (as they figure it). (See our full statement about data privacy, attached to this registration, and also accessible here: https://colorgame.net/downloadables/data-privacy-full.pdf.)

In order to attract the widest possible number of players, the app was translated into four languages (plus English): Chinese, French, Spanish, and German. It will be launched in two steps: a technical launch will be followed by a seeding period where the app will be restricted to a group of guests recruited through social networks (and excluding anyone connected in any way with the app's development); two weeks into this period, the app will be officially launched and opened to the broader public. We set up a mailing contact and a "subreddit" forum to address possible questions and conversations from the players. Since we'll be moderating this forum, we can make sure that players do not use it to coordinate around shared codes outside the game.

C. Registered hypotheses

We have prepared (sometimes in collaboration with external researchers) six studies containing predictions about the game. They have been preregistered on the open science framework together with this document to establish a sharp distinction between true predictions and post-dictions. While this does not prevent exploratory analyses, it implies that such analyses will be clearly signalled as such and treated separately from actual hypothesis-testing. The complete data gathered from the app will be made public to other researchers who may then exploit it in their own ways.

Since we do not want the app's users to be aware of our expectations, the hypotheses and methods for our six studies are not yet public. They are listed on the Open Science Framework with nicknames that we also use for cross-referencing in the registration documents: FRIENDS, INFORMATION, LANGUAGE, PRIORS, SALIENCE, TREES. At the end of the one-year data collection period, the registration documents, a time-stamped record of our hypotheses and methods, will be released automatically on the Open Science Framework. We will also keep there an online diary of the app’s functioning, to document major events (or possible changes) in the game.

The specifics of data selection and analysis are registered separately for each study, but there are a few constraints that apply to all studies.

The data collection period runs for one year from the official launch of the game. After this period has passed, the data collected by the app will no longer be taken into account for the purpose of testing our original preregistered hypotheses. The app may no longer be maintained or advertised, although it will still be possible to download it for a period of time, and it will keep recording data.

The following players and trials will be excluded from the data:
- IP addresses that show cues of bot behaviour: for instance, if an IP plays an unnaturally high number of trials per day; or if a number of machines enter the game at the same hour from one single city, in a country known to host anomalous behaviour on the cyberspace, each machine playing one or a few trials. Such incidents will be signalled in the diary.
- Trials where Sender did not send any symbol will be excluded (this may happen in speed mode).
- All players have the legal right to request the removal of their data from our dataset without giving a reason (see the document on data protection attached to the registration,

also here: https://colorgame.net/downloadables/data-privacy-full.pdf).
- Our technical report, finalised after the end of the one-year period, may uncover sources of corrupt data that warrant further exclusion measures. These exclusions will be decided on a technical basis and will apply to all six studies (regardless of what they predict).

After the end of the one-year testing period, a technical report will be released to check for the quality of the data obtained by the app. This report will not be testing any hypothesis, but will consist in a series of quality controls. Among other things, it will address the following questions:

- Do individual player identifiers display patterns of use consistent with a single user? The Color Game does not require players to log in individually in order to play the game. This allows several players to use the same smartphone. Most of our studies are interested in data at the level of trials or on the scale of the game as a whole, so this will not be a problem, but we still want to be reasonably sure that most machines represent one player (not a dozen players). This will also be an occasion to filter out possible bots. We will check the number of trials per day and test whether most machines become better at the game with time.

- Is colour perception markedly biased by colour modulation programs for smartphone screens? Some smartphone built-in functionalities dim the light or shift colours towards the red spectrum after dusk. To make sure that this does not disrupt the data in ways that would be too detrimental to our studies, we will check that success rates do not change depending on the hour of the day (using the users' language preferences as a way to infer their timezone).

This list of possible checks is open-ended.

D. Hypothesis testing: The app vs. the lab

Most of our hypotheses, but not all of them, bear on randomised interventions inside the app: the random allocation of players to groups and sub-groups, the random disclosure of colours to Senders. While randomisation, coupled with a very large number of participants, rules out confounding factors that may bias a non-randomised test, it still does not provide us with the kind of controlled environment we could have obtained in the lab. We cannot, for instance, be certain that all of our participants will have normal colour vision or that some smartphone screens will not be calibrated in ways that significantly bias colour perception. The app, however, also greatly increases our tests' statistical power, compared to a standard experiment, due to the large potential number of participants and observations. Under-powered studies are a pervasive and urgent problem in experimental research (Bakker, Dijk, and Wicherts 2012; Maxwell 2004). A limited relaxation of our usual level of control over participant behaviour is, we believe, a decent price to pay to tackle this issue.

We also plan to test hypotheses without relying on randomised interventions, observing correlational patterns and evolutionary trends whose study is crucial to language evolution research (Gray, Drummond, and Greenhill 2009; Pagel, Atkinson, and Meade 2007). The app's data will improve upon standard historical data in several ways. The app should provide an exhaustive record of all symbol uses within the game, thus doing away with the recurrent problem of inferring the state of a language from limited or biased sources. Another improvement comes from the random allocation of players to two independent halves. Cultural change is a contingent and path-dependent process (Salganik and Watts 2009): Would we still observe the same patterns if the tape were replayed, so to speak? (To borrow a metaphor from Gould, 2000.) Historical data provide semi-independent trajectories that can be brought to bear on this question, but our two-halves system provides a more controlled way to duplicate such observations. Lastly, the app's data is fully structured: no classificatory choices will need to be made after data collection. Both the reference space (the 35 colours) and the symbols are identified unambiguously, avoiding the categorisation issues that may affect natural language inventories (Haspelmath 2007).

The app can be used to test methods in addition to hypotheses. Phylogenetic algorithms are a case in point. Descent relationships among languages can be inferred, to a certain extent, using a variety of statistical tools borrowed from evolutionary biology (Dunn 2014). This method has generated much attention and excitement, but is yet to become fully consensual. Inferences concerning past events can never be fully validated, and phylogenetic methods are no exception. This has caused a great deal of debate among students of cultural and linguistic evolution (e.g. Bouckaert et al. 2012; Croft 2001; Greenhill, Currie, and Gray 2009). Biologists have found a way around this problem with experimental phylogenetics (Hillis et al. 1992; Rozen, Schneider, and Lenski 2005), which cultural evolution has started to emulate (Schillinger, Mesoudi, and Lycett 2016). Experimental phylogenies simulate a process of repeated genetic or cultural transmission, which is thus perfectly known, and tests phylogenetic inference algorithms by running them on the data so produced, comparing their output with reality. Yet the inherent time and space constraints of laboratory experiments here again prove limiting. Biologists overcome them by using organisms with atypically short lifecycles and artificially boosted mutation rates, but this leads to biases (Oakley 2009). Artificial cultural transmission chains are also problematic (see section A). Still the most important limitation of experimental methods is here again one of scale, and this probably explains why no experimental validation of language phylogenies has been tried yet. The app's free grouping and merging tool allows it to simulate the population-splitting and population-merging events that play a fundamental role in the emergence of languages, dialects and creoles (Croft 2001).

E. Conclusion

Linguists have used data from pre-existing online games before (Skirgård, Roberts, & Yencken, 2017), but building a gaming app for specifically scientific purposes is new. Although we have been focusing on artificial language experiments, the range of application for smartphone apps is much broader. Language documentation provides obvious opportunities: many linguistic questionnaires can already be automated and self-administered, giving participants greater control over their content. Future developments in natural language recognition may be quite helpful here. Compared to crowdsourcing of a standard experiment, apps are costly and may be difficult to create, not least because the public demands to be engaged and entertained in exchange for their participation (unlike paid experimental participants). Yet the recent successes of various "serious games" and "citizen science" initiatives (Cooper et al. 2010) show this is doable. We hope to have shown in this paper that one need not sacrifice the methodological demands of experimentation and hypothesis testing when dealing with large app-generated datasets.

Acknowledgements: Our warmest thanks go to all the collaborators and volunteers who have helped us develop and promote the game over the last two years: Follow the URL http://colorgame.net/downloadables/acknowledgements.pdf to know who they are.

Appendices: This presentation paper is part of a registration folder that includes other documents, listed on the Open Science Framework page.

References

Bakker, Marjan, Annette van Dijk, and Jelte M. Wicherts. 2012. “The Rules of the Game Called Psychological Science.” *Perspectives on Psychological Science* 7 (6): 543–54. https://doi.org/10.1177/1745691612459060.

Bartlett, Frederic. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Baumard, Nicolas, Jean-Baptiste André, and Dan Sperber. 2013. “A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice.” *Behavioral and Brain Sciences* 36 (1): 59–78. https://doi.org/10.1017/S0140525X11002202.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 2012. “Mapping the Origins and Expansion of the Indo-European Language Family.” *Science* 337 (6097): 957–60. https://doi.org/10.1126/science.1219669.

Brown, Harriet R., Peter Zeidman, Peter Smittenaar, Rick A. Adams, Fiona McNab, Robb B. Rutledge, and Raymond J. Dolan. 2014. “Crowdsourcing for Cognitive Science – The Utility of Smartphones.” *PLOS ONE* 9 (7): e100662. https://doi.org/10.1371/journal.pone.0100662.

Casler, Krista, Lydia Bickel, and Elizabeth Hackett. 2013. “Separate but Equal? A Comparison of Participants and Data Gathered via Amazon’s MTurk, Social Media, and Face-to-Face Behavioral Testing.” *Computers in Human Behavior* 29 (6): 2156–60. https://doi.org/10.1016/j.chb.2013.05.009.

Claidière, Nicolas, and Dan Sperber. 2010. “Imitation Explains the Propagation, Not the Stability of Animal Culture.” *Proceedings of the Royal Society B*.

Clark, Herbert H., and Deanna Wilkes-Gibbs. 1986. “Referring as a Collaborative Process.” *Cognition* 22 (1): 1–39. https://doi.org/10.1016/0010-0277(86)90010-7.

Cooper, Seth, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. 2010. “Predicting Protein Structures with a Multiplayer Online Game.” *Nature* 466 (7307): 756–60. https://doi.org/10.1038/nature09304.

Croft, William. 2001. *Explaining Language Change*. 1st ed. Pearson Education ESL.

Dingemanse, Mark, Seán G. Roberts, Julija Baranova, Joe Blythe, Paul Drew, Simeon Floyd, Rosa S. Gisladottir, et al. 2015. “Universal Principles in the Repair of Communication Problems.” *PLOS ONE* 10 (9): e0136100. https://doi.org/10.1371/journal.pone.0136100.

Dunn, Michael. 2014. “Language Phylogenies.” In *The Routledge Handbook of Historical Linguistics*, edited by Claire Bowern and Bethwyn Evans, 190–211. London: Routlege.

Enfield, N. J. 2017. *How We Talk: The Inner Workings of Conversation*. New York: Basic Books.

Fitzpatrick, Kathleen Kara, Alison Darcy, and Molly Vierhile. 2017. “Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial.” *JMIR Mental Health* 4 (2). https://doi.org/10.2196/mental.7785.

Gandomi, Amir, and Murtaza Haider. 2015. “Beyond the Hype: Big Data Concepts, Methods, and Analytics.” *International Journal of Information Management* 35 (2): 137–44. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Gould, Stephen Jay. 2000. *Wonderful Life: Burgess Shale and the Nature of History*. New Ed. London: Vintage.

Gray, R. D., A. J. Drummond, and S. J. Greenhill. 2009. “Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement.” *Science* 323 (5913): 479–83. https://doi.org/10.1126/science.1166858.

Greenhill, Simon J., Thomas E. Currie, and Russell D. Gray. 2009. “Does Horizontal Transmission Invalidate Cultural Phylogenies?” *Proceedings of the Royal Society* *of London B: Biological Sciences* 276 (1665): 2299–2306. https://doi.org/10.1098/rspb.2008.1944.

Griffiths, Thomas L. 2015. "Manifesto for a New (computational) Cognitive Revolution." *Cognition* 135 (February): 21–23. https://doi.org/10.1016/j.cognition.2014.11.026.

Haspelmath, Martin. 2007. "Pre-Established Categories Don't Exist: Consequences for Language Description and Typology." *Linguistic Typology* 11 (1): 119–32. https://doi.org/10.1515/LINGTY.2007.011.

Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineux. 1992. "Experimental Phylogenetics: Generation of a Known Phylogeny." *Science* 255 (5044): 589–92. https://doi.org/10.1126/science.1736360.

Kirby, Simon, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. "Compression and Communication in the Cultural Evolution of Linguistic Structure." *Cognition* 141 (August): 87–102. https://doi.org/10.1016/j.cognition.2015.03.016.

Klindt, David, Marie Devaine, and Jean Daunizeau. 2017. "Does the Way We Read Others' Mind Change over the Lifespan? Insights from a Massive Web Poll of Cognitive Skills from Childhood to Late Adulthood." *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 86: 205–15. https://doi.org/10.1016/j.cortex.2016.09.009.

Lathia, Neal, Gillian M. Sandstrom, Cecilia Mascolo, and Peter J. Rentfrow. 2017. "Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity." *PloS One* 12 (1): e0160589. https://doi.org/10.1371/journal.pone.0160589.

Luo, M. R., G. Cui, and B. Rigg. 2001. "The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000." *Color Research & Application* 26 (5): 340–50. https://doi.org/10.1002/col.1049.

Maxwell, Scott E. 2004. "The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies." *Psychological Methods* 9 (2): 147–63. https://doi.org/10.1037/1082-989X.9.2.147.

Mesoudi, Alex, and Andrew Whiten. 2008. "The Multiple Roles of Cultural Transmission Experiments in Understanding Human Cultural Evolution." *Phil. Trans. R. Soc. B* 363: 3489–3501.

Miller, Geoffrey. 2012. "The Smartphone Psychology Manifesto." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 7 (3): 221–37. https://doi.org/10.1177/1745691612441215.

Morin, Olivier. 2015. *How Traditions Live and Die*. Cambridge MA: Oxford University Press.

Mueller, Thomas, James Winters, and Olivier Morin. n.d. "How Does Shared Information Influence the Evolution of Graphic Codes?"

Noë, Ronald, and Peter Hammerstein. 1994. "Biological Markets: Supply and Demand Determine the Effect of Partner Choice in Cooperation, Mutualism and Mating." *Behavioral Ecology and Sociobiology* 35 (1): 1–11. https://doi.org/10.1007/BF00167053.

Oakley, Todd. 2009. "A Critique of Experimental Phylogenetics." In *Experimental Evolution*, edited by Garland Theodore and Rose Michael, 659–71. Los Angeles: University of California Press.

Pagel, Mark, Quentin Atkinson, and A. Meade. 2007. “Frequency of Word-Use Predicts Rates of Lexical Evolution throughout Indo-European History.” *Science* 449: 717–21.

Rozen, Daniel E., Dominique Schneider, and Richard E. Lenski. 2005. “Long-Term Experimental Evolution in <Emphasis Type=‘Italic’>Escherichia coli</Emphasis>. XIII. Phylogenetic History of a Balanced Polymorphism.” *Journal of Molecular Evolution* 61 (2): 171–80. https://doi.org/10.1007/s00239-004-0322-2.

Salganik, Matthew J., and Duncan J. Watts. 2009. “Web-Based Experiments for the Study of Collective Dynamics in Cultural Markets.” *Topics in Cognitive Science* 1: 439–68.

Schillinger, Kerstin, Alex Mesoudi, and Stephen J. Lycett. 2016. “Copying Error, Evolution, and Phylogenetic Signal in Artifactual Traditions: An Experimental Approach Using ‘model Artifacts.’” *Journal of Archaeological Science* 70 (June): 23–34. https://doi.org/10.1016/j.jas.2016.04.013.

Scott-Phillips, Thomas C., and Simon Kirby. 2010. “Language Evolution in the Laboratory.” *Trends in Cognitive Sciences* 14 (9): 411–17. https://doi.org/10.1016/j.tics.2010.06.006.

Skirgård, Hedvig, Seán G. Roberts, and Lars Yencken. 2017. “Why Are Some Languages Confused for Others? Investigating Data from the Great Language Game.” *PLOS ONE* 12 (4): e0165934. https://doi.org/10.1371/journal.pone.0165934.

Yule, George. 2013. *Referential Communication Tasks*. Routledge.