An overview of the "Color Game" App project

Olivier Morin[1,2], James Winters[3], Thomas Müller[4], Tiffany Morisseau[5,6]

[1] Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, UMR 8129. 29, rue d'Ulm, 75014 Paris.
[2] Max Planck Institute for the Science of Human History, Department of Linguistic and Cultural Evolution, Jena, Germany.
[3] UM6P School of Collective Intelligence, 43150 Ben Guerir, Morocco.
[4] Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany.
[5] Université de Paris, LAPEA, Boulogne-Billancourt, France
[6] LAPEA, Université Gustave Eiffel, IFSTTAR, Versailles, France

Corresponding author: Olivier Morin, morin@shh.mpg.de.

The Color Game gaming app (2018–2019) invited players from all over the world to invent a visual language without words. Participants took part in a referential communication task where a Sender had to indicate a colour to a Receiver, with the help of black and white symbols. They could freely choose which other players they interacted with, and play repeatedly with their chosen contacts. This paper presents the Color Game dataset, accessible at https://osf.io/9yc25/, which records all interactions between app players. In its final cleaned-up version, the dataset contains 347,606 games by 2,535 players, from more than 100 different countries, speaking 80 different languages. This companion paper describes the app's workings and history.

1. General description

The Color Game app was a gaming app designed to investigate the creation of an artificial language on 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 (*sensu* Yule 2013). The task was piloted in the lab and used to generate results published in (Müller, Winters, and Morin 2019) before we turned it into an online game.

The task involves two players. The first player (the Sender) is presented with a "target colour", and his goal is to communicate with the second player (the Receiver) to help them pick the target colour in an array of four, thereby earning points. Senders 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 Color Game 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, Winters, and Morin 2019).

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 960 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.

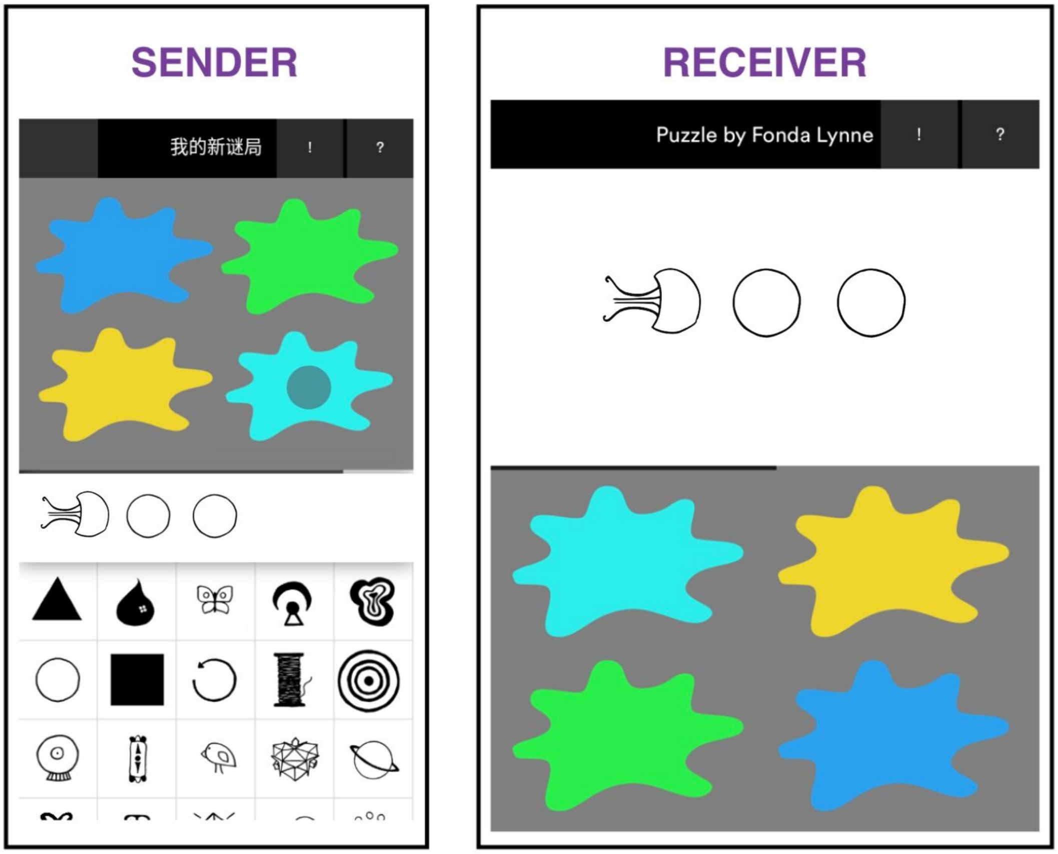


Fig. 1. A trial in synchronous mode. The Sender (her screen is shown on the left panel) communicates with the Receiver (right panel) to help the Receiver find the target colour (here, the lighter shade of blue), marked (for Receiver) by a dot. The Sender selects symbols from her keyboard on the bottom of the screen. The symbols are arranged in random order; scrolling down would reveal additional symbols. The symbols tapped by the Sender show in the rectangle window in the middle. They also show instantly on the Receiver's screen. In the black bar on top, both players can see the identity of the Sender or Receiver in the language they chose, as well as two symbols, "!" and "?". Either player can tap of one of these symbols. This will highlight and enlarge the symbol for both players and produce a sound. In asynchronous mode, the set-up is similar, but the "!" and "?" symbols are not available.

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 the corresponding points. After every block, the Receiver is told how many of the 10 they got correct, but not given specific information about which ones. 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 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.

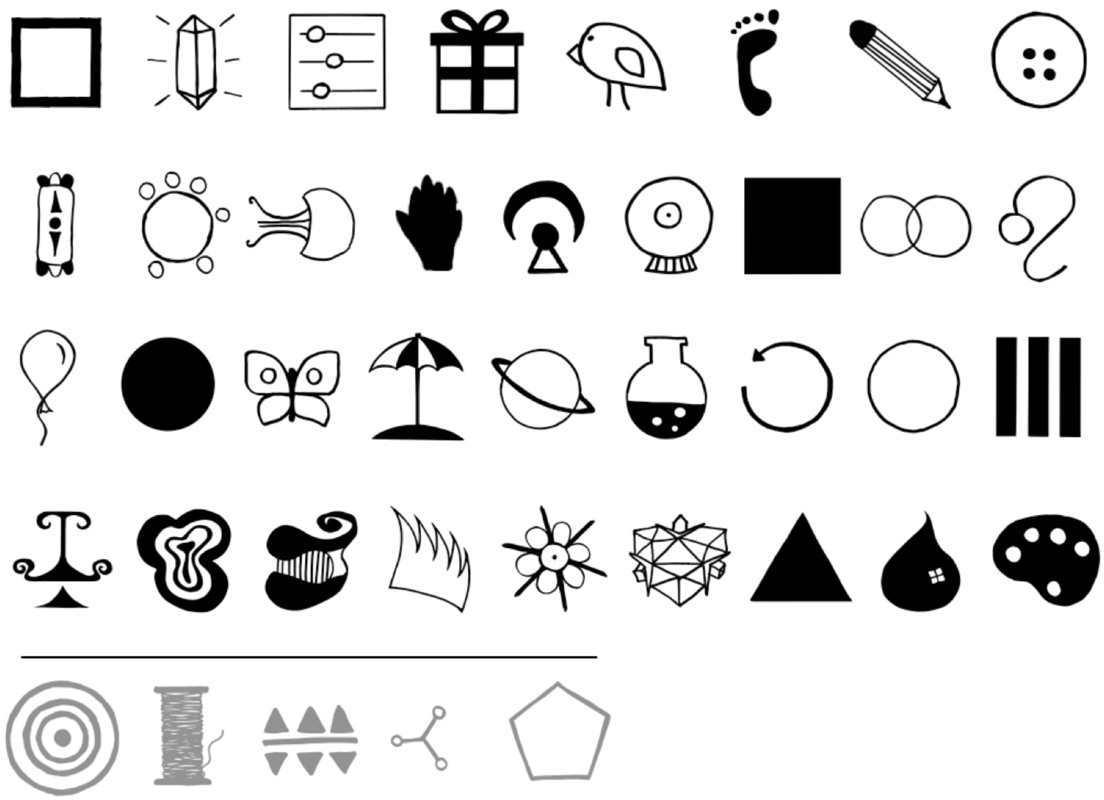


Fig. 2. The 35 symbols used in the game (first four rows). Bottom row, in grey: the five symbols used for the tutorial and for the videos advertising the game (these symbols are for tutorials only). The tutorial symbols were also used in the advertisement video we released to promote the game.

The game's colour space was 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 were 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 arrays of four colours were formed from this set of 32 colours, by picking every fourth colour along the dimension of hue, using each of the 32 colours as starting point (**Fig. 4**.). Each colour thus appeared in four different arrays. The arrays were 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 could see some or all of the colours visible in the Receiver's array. This quantity varied from one (only the target) to four (the full array), and was randomised.

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

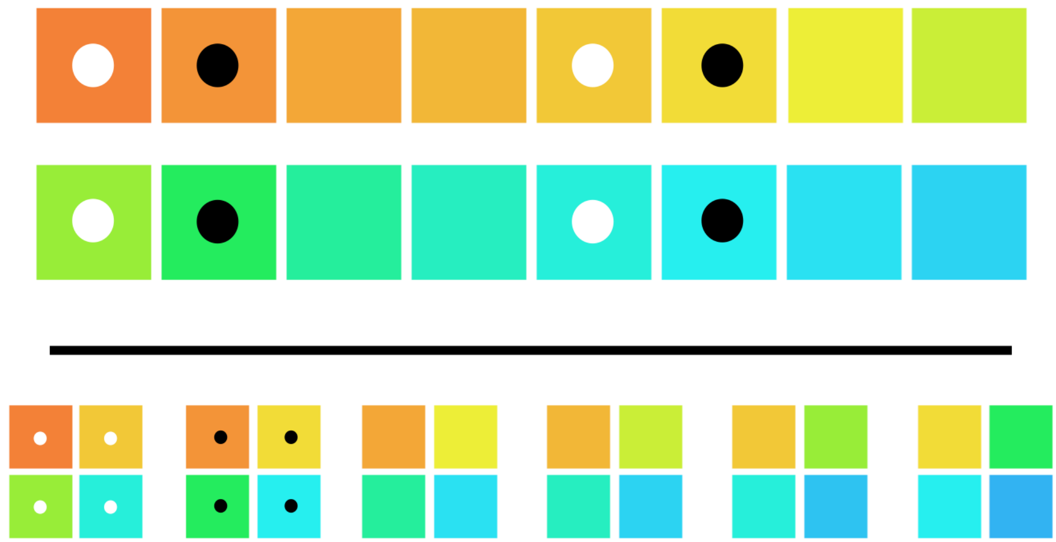


Fig 4. How colour arrays were built. Top row: The composition of two colour arrays, one marked by white dots, the other by black dots, is shown relative to the colour space. Bottom row: Six contiguous colour arrays (out of 32), including the white-dot and black-dot ones.

Every new player, on their first opening of the app (but not later) was greeted with a short tutorial explaining the basics of the game. The tutorial simulated a referential communication game, using dummy symbols that would never be re-used in the normal course of the game (**Fig. 2**, bottom row), and a colour array randomly picked from the 32 possible ones. The player was presented with a symbol and an array, and asked to point the colour it could refer to. Next, the player was asked to play as Sender and to use one of the dummy symbols to refer to a target colour. They were then told they had completed that step of the tutorial. After this introduction, the players were given a guided tour of the home screen, with pop-up messages in their chosen language.

The app left players free to choose their interlocutors and the format of their interaction. It allowed for both "synchronous" (i.e., live, or in real time) and "asynchronous" play. Senders playing asynchronously simply typed black and white symbols corresponding to the target colour. The symbols were then sent to Receivers along with the colour array, and Receivers figured out the target from these symbols. Their message was always accompanied by the corresponding array of four colours, out of which the Receiver must pick the target. These *asynchronous puzzles* remained available inside the app indefinitely as long as no one had played them, but each puzzle disappeared as soon as one Receiver had played it. When playing synchronously, both players received feedback on their performance. When playing asynchronously, the Receiver received similar feedback, but the Sender was simply notified that someone had played with their messages, earning them the corresponding points.

Synchronous play required the two players to contact each other and stay connected for as long as they played. This enabled them to communicate in real time, and exchange repair signals consisting of the signs “?” and “!”. The players were not told what these punctuation marks precisely meant in the context of the game, and we expected variation in the way they were used. Asynchronous play made 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 could observe, while synchronous play allowed us to study interactional properties of communication, such as repair. A “notebook” feature allowed players to write down the meaning of particular symbols (as they figured it).

The app’s players were free to choose their partners from a vast pool of players (**Fig. 5**). To play with another contact, a player either invited them for synchronous play and waited for the invitation to be accepted, or sent an asynchronous message, which could be broadcast to the whole group or sent to a specific individual contact who could open it at a later time. Players were incentivized to play as Senders, so that the app would always contain a sufficient number of high-quality asynchronous puzzles. Receivers had to pay a number of points to the Sender whose puzzle they wanted to play, and Senders received these points regardless of the Receiver's performance. Extra points could be earned by playing an especially difficult, time-limited “speed mode”, available to players after a certain level.

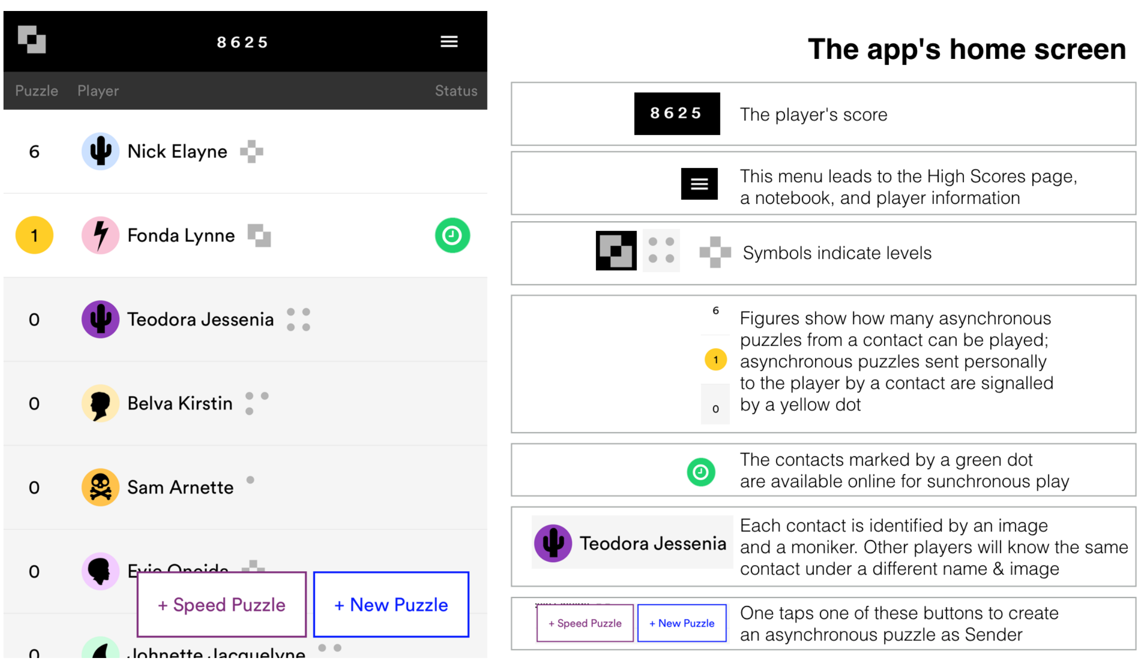


Fig. 5. The app's home screen (left) with legend (right). The colourful logos that identify each contact are randomly generated from a set of black and white pictures and a set of colours.

We expected to see the emergence of a market dynamic, whereby the most skilful players would get more proposals for synchronous play and more takers for their asynchronous puzzles. The app made 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, was meant to discourage the creation of low-quality puzzles or the absorption of high-quality puzzles by negligent players.

Players reached new levels as they earn more points (**Table 1**). In addition to being symbolic rewards, higher levels allowed players to add new symbols to their keyboard. The allocation of symbols was always random: a beginner would start at level 0 with 15 randomly assigned symbols, rising gradually to a full keyboard of 35.

Level	**Number of points**	**Number of symbols**
0	0	15
1	250	17
2	500	19
3	750	21
4	1250	23
5	1500	27
6	2000	31
7	2500	33
8	3000	35
9	5000	35
10	7500	35
11	15000	35
12	40000	35
13	80000	35

Table 1: Color Game levels. The levels achieved by players of the Color Game, with the number of points required to ascend to each level and the number of symbols that each level gave access to the player (symbols that the player could use when playing as Sender). Level 10 was notable since players at that level and above could choose to play as Senders in stress mode.

Strict player anonymity was ensured by a system of pseudonyms and “cover names”. The app recorded no personal identifiers (such as names), but gave each player a public pseudonym, displayed on the “top players” score board. Crucially, however, that pseudonym could not be used to identify a player as a contact to play with. Instead, contacts were only ever known to one another by means of “cover names” that varied from contact to contact and from player to player. Player X knew her contact through a list of randomly generated names that reliably identified contact A, contact B, contact C, and so on. Player Y could also identify the same contacts A, B, C, but by completely different names, so that A’s name for X did not correspond in any way to A’s name for Y, even though both players could reliably identify and contact A. This ensured that players could not circumvent the app to contact one another in real life, on social networks, etc.

Players were divided, when the app was launched, between two pools of players called “Halves”, with each new player being assigned randomly to one half. All interactions between players were restricted to their half, and this primary division stood in place indefinitely. We planned to further divide our pool of players into more plastic sub-groupings, to create population divergences or merging events, an important condition of one of our projects (TREES). However, the pool of players was not big enough to make this a safe option, and this project was not carried out.

Participants agreed to have their data collected in an anonymous format and for research purposes in a consent form they filled out 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). (See our full statement about data privacy, section 8). In addition to recording all the players’ moves as Sender or Receiver, the app also recorded notes taken by the players who made use of “the notebook”.

In order to attract the widest possible number of players, the app was translated into four languages (plus English): Chinese, French, Spanish, and German. Later on (September 2018), we added Portuguese (Brazil), Japanese, and Russian (see our deployment log, section 4).

Following an extensive period of alpha- and beta-testing (all beta players involved in the piloting and debugging, most of them friends and collaborators of the authors and developers, were barred from taking part in subsequent stages), the App was launched in two steps. A technical launch (April 26th, 2018) opened a “seeding period” where the app was restricted to a group of guests recruited through social networks (and excluding anyone connected in any way with the app’s development). 80 players were thus brought in. They allowed us to fine-tune the app's points system, and seeded the app with puzzles that new players would be able to solve. Two weeks later (May 15th), the app was officially launched and opened to the broader public through three main media: the Google Play store, the App Store (for iPhone), and the website colorgame.net. We set up a mailing contact and a "subreddit" forum to address possible questions and conversations from the players. Since we were moderating this forum, we could make sure that players would not use it to coordinate around shared codes outside the game. We also kept a Twitter account. The deployment log in section 4 below provides a complete chronology of the app's life cycle.

2. Preregistered predictions & projects

2.1. Preregistration process

The Color Game project and the six studies that constitute it were preregistered on the Open Science Framework on April 11th, 2018. The full registration can be found here: https://osf.io/9pdzk. It includes the source code for the app, a plan for data collection along with criteria for including and excluding data points, and the specific predictions for each of the six projects. Since we did not want the app's users to be aware of our expectations, the hypotheses and methods for our six studies were not made public when the app was launched. They were 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 time the registration took place, we did not communicate with the public about our particular project, but did a presentation of the Color Game as a whole in (Morin et al. 2018).

2.2. The projects

For each project we provide a link to a public OSF repository, which contains

- The project's preregistration (or preregistrations) (under "Registrations");
- A link to a preprint of the paper taken from the project, the preprint itself, or a report (except for the PRIORS project, still in progress);
- The data and code needed to replicate the preregistered analyses.

2.2.1. FRIENDS (https://osf.io/y2vak/).

(Authors: Morin, Müller, Morisseau, Winters.)

This project tested the hypothesis that the precision of symbol-colour mappings would increase through time. Symbols' meanings were defined as the distribution of colours a given symbol was associated with for each Sender. We measured the precision of symbol meaning using entropy (more precisely, a circular measure of entropy specifically designed for this project by Winters). Our prediction was validated, although not in its original preregistered version. This project became a paper that is currently submitted to a journal.

2.2.2. INFORMATION (https://osf.io/7y9pn/).

(Authors: Morin, Winters, Müller, Morisseau.)

This project aimed to find out whether the amount of information that a Sender possessed concerning the colour array seen by the Receiver would influence performance in referential communication. Starting from the fact that the number of colours seen by Senders randomly varied between 1 and 4, we calculated the amount of information that each colour shown to Sender revealed about the full array. We hypothesized that this quantity of information would predict performance better than the absolute number of colours seen by Senders. This hypothesis was verified. However, we also found that this effect was crucially modulated by the position of the colour target inside the array (whether it was peripheral or central), something that our model did not predict. This project became a report, not currently submitted to a journal.

2.2.3. LANGUAGE (https://osf.io/a8bge/).

(Authors: Müller, Winters, Morisseau, Noveck, Morin.)

The goal of this project was to investigate a possible impact of the colour categories present in the players' native language on the categorization of colours through symbol use within the Color Game. Using data from a separate survey on colour naming, we determined the semantic structure of three languages with respect to colours. Confirming our predictions, native and artificial language structure overlap at least moderately. Communicative behaviour and performance were influenced by the shared semantic structure, but only for English-speaking pairs. This project became a paper that is currently submitted to a journal.

2.2.4. PRIORS (https://osf.io/dqhtv/).

(Authors: Winters, Müller, Morin.)

This project investigated conventions and biases in the emergence and evolution of symbol-colour mappings. Using a separate task in which participants were asked to associate one of the game's 35 symbols with each of its 32 colours, we were able to assess "prior" values for the probability of associating a colour with a symbol, independently of any communicative interactions. Our predictions concerned the influence of these prior associations over the conventional associations that arose in the Color Game. Analyses for this project are still in progress.

2.2.5. SALIENCE (https://osf.io/f9xzq/).

(Authors: Morin, Müller, Morisseau, Winters.)

This project tested the prediction that focal colours (focal meaning here, centrally associated with basic colour terms) are easier to communicate. The association of focal colours with basic colour terms in five languages was asserted with an online survey independent of the app. We predicted that focal colours would be easier to communicate, and that codes indicating focal colours would be more stable through time. These predictions were validated. This project became a paper that is currently submitted to a journal.

2.2.6. TREES (https://osf.io/r7n32/).

(Authors: Morin, Morisseau, Winters, Müller, Greenhill.)

The goal of this project was to simulate the evolution of distinct dialects inside the Color Game, and to retrace their emergence with a phylogenetic algorithm. A necessary first step in this project was to "grow" different dialects inside the Color Game. To this end, we had separated the game's players into two "Halves". The app did not allow players from one Half to play with a player identified by their device as belonging to the other Half. We expected that distinct dialects would emerge in each Half, and that different symbols would be paired with different colours. We observed dialectal differences for a small minority of symbols (between 2 and 5 out of 35). Even for these symbols, the differences are subtle. Thus, dialectal differences were not important enough to make the TREES project viable. Carrying out to completion would have involved dividing the two Halves into sub-groups, with the risk of losing too many players and no guarantee of exploitable data. A report on the preliminary results was uploaded.

2.3. Open-ended exploration

In addition to the preregistered analyses, we also carried out some non-preregistered descriptive analyses, to see how well players succeeded, when and with whom they played, etc. These analyses were exhaustively recorded in our research diary, reproduced in section 5.

3. Open data & code

3.1. The Color Game dataset repository

The open Color Game dataset is available on the Open Science Framework (https://osf.io/9yc25/). This page contains all the information necessary to explore, use, and cite the data. We provide both the raw data as outputted by the app, and the cleaned-up dataset which we used for all the analyses involved in our six projects (the exact same dataset was used systematically). We provide the code that we used to get from the raw version to the cleaned-up one, and a comprehensive description of all the deletions we made, and why they were made.

3.2. Exclusion and inclusion criteria: preregistered rules

We preregistered a number of rules to create, select, or exclude data points. They are copied below:

"*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)."

3.3. Exclusion and inclusion criteria: departures from the preregistered rules

The main departure from our preregistered rules was that we ended the App's lifespan a month earlier than planned (April 10^{th}, 2019). Activity had slowed down considerably at that point (see our deployment log, section 4). Other departures were motivated by unplanned bugs (see the report on the Clean-up script for a full description of exclusions).

3.4. Other datasets

In addition to the main dataset, we constituted two other independent datasets to be used for specific projects (PRIORS, SALIENCY, and LANGUAGE). These tasks were carried out online with completely independent pools of subjects recruited through Prolific or SurveyMonkey. The PRIORS project used a task that asked participants to pair colours and symbols according to their intuitions, independently of any communicative interactions, to obtain data on prior associations between symbols and colours. The SALIENCY and LANGUAGE projects both made use of data from an online experiment whose participants had to name colours in their native language. These datasets are available on the page of their respective projects.

3.5. Open code

The source code for the Color Game app is available on GitHub (https://github.com/ColorGameMPG/ColorGame) and copied on the project's main OSF page (https://osf.io/wq3uz/). The R scripts used to analyse the data for various projects is uploaded on these projects' respective OSF pages. The R script used to clean up the data is provided on the Color Game dataset OSF page.

4. Color Game deployment log

We kept on the Open Science Framework an online diary of the app's functioning, to document major events (or possible changes) in the game. It is reproduced here. It is also accessible at this address: https://osf.io/eaxkj/

26.04.2018: The game's "soft launch" has started. Between now and the 15th of May, we will invite around 100 journalists and friends to "seed" the game and explore it before the full public launch. Anyone who has already played the Color Game has received a message discouraging them from playing this version.
27.04.2018: Thomas Müller & James Winters promote the game at a games conference in Berlin, allowing visitors to test the app.
27.04.2018: A press release was sent to around 50 journalists (English and French-speaking mostly). It included an invitation to access the app before its launch.
04.05.2018: The "soft launch" has brought around 80 players to download the app. I (OM) have been regularly tuning the game's parameters. [The] threshold of successful trials

necessary to win a puzzle set has been changed from 4 to 3; the cost of puzzles has been changed several times, as well as the points gained by players after playing them. The current settings for the game are uploaded as “settings-2018-05-04.jpg” in the OSF repository. [**Fig. 6**] The current settings for the Levels is also uploaded: “levels-2018-05-04.csv”. [Identical to Table 1 above.]

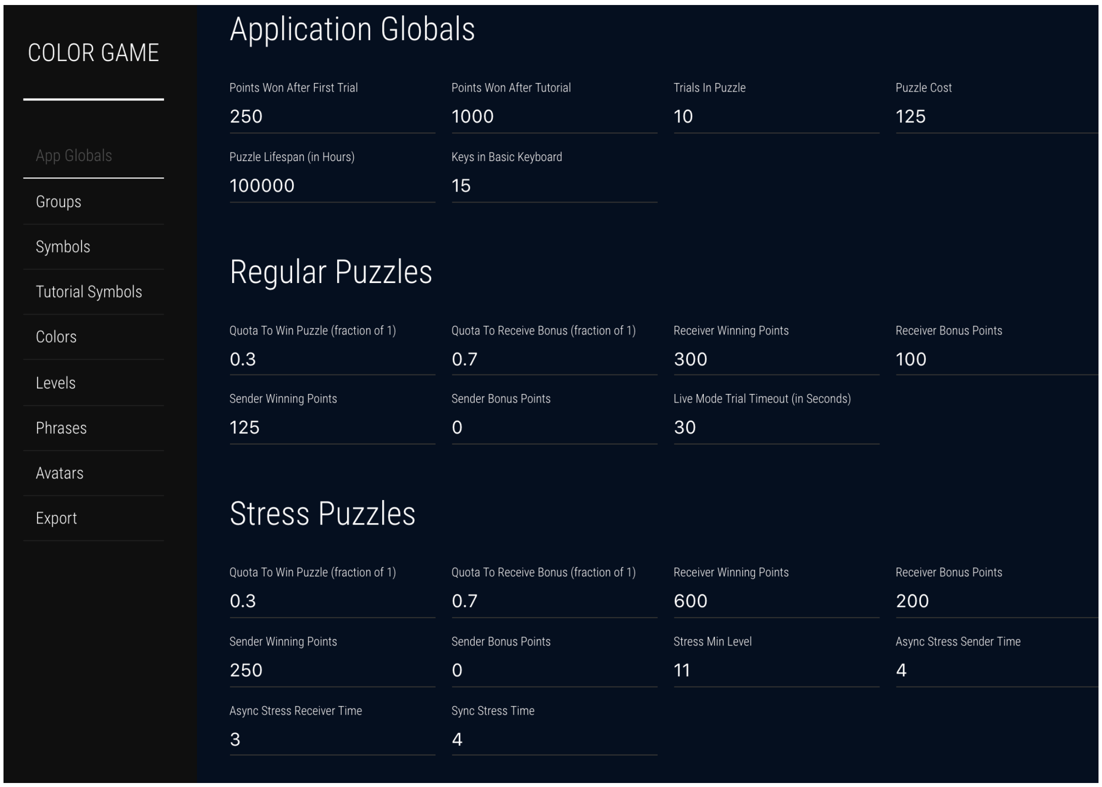


Fig. 6. The Color Game’s parameters on May 4th, 2018. (Screen capture of the app’s back end.)

15.05.2018: The game’s public launch took place during the night, the embargo for our press release expiring at midnight CET. Last Sunday I fine-tuned the points system one last time (number of points players get at the outset). The levels system is unchanged. I upload the current settings as settings-2018-05-15.jpg. [**Fig. 7.**]

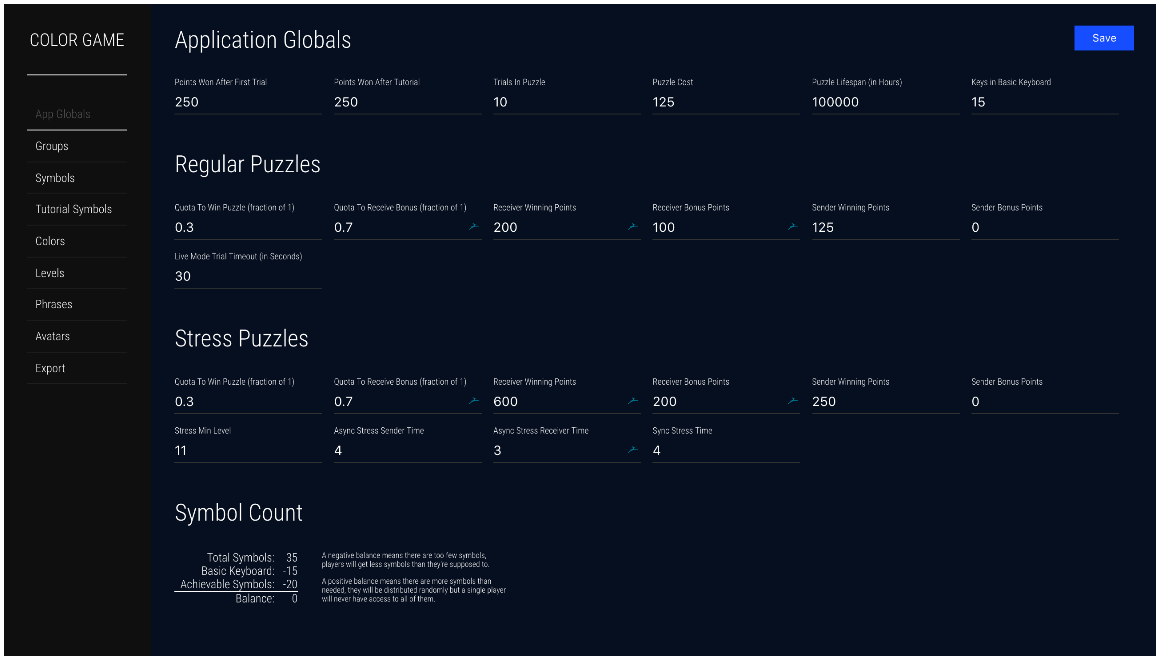


Fig. 7. The Color Game's parameters on May 15th, 2018 (Start of the public launch and official data collection period), and subsequently. (Screen capture of the app's back end.)

21.05.2018: The affluence of players this week proved taxing for our system. Smartphones need to refresh regularly a set of several hundred profiles, which occasions delays. To avoid this, as a temporary fix, we decided to split the two Halves, Half A and Half B, into two sub-halves (A1, A2 from A, B1, B2 from B). The two splits took place at 9:12 AM CET (recorded as 7:12 in the data, which gives time as GMT). This is intended as a temporary fix: the sub-groups will be merged back as soon as this is doable without running into our current problems. The developers have been instructed to erase all the profiles of users that have not completed the tutorial, to lighten the load. More long-term solutions are being studied.

22.05. 2018: The app has now been installed over 2000 times. A little less than 600 players have reached level 2 (Level 1 is reached automatically upon completing the tutorial). Approximately 1000 puzzles (=10 000 trials) have been played each day. Players are rather diverse and come from more than 60 different countries.

29.05.2018: To fix the issue that was identified on 22.05 (see above), the developers released a new version of the app, where each player only sees other players if they (a) — have puzzles available for that player, and/or (b) have interacted with that player, and/or (c) are live. This restriction avoids the proliferation of players that plagued the app and caused bugs. The updated version was released on the 28th for Android (morning) and iOS (evening). Android versions of the Color Game were automatically updated, while iOS versions had to be redownloaded to get the update. Today at 13:13 CET (11:13 GMT), the

2x2 groups that had been split were re-merged into the moieties that they came from. Groups A1 and A2 were re-merged into Half A, and the same happened to B1 and B2.
20.08.2018: A player made us aware that he was able to solve the same puzzle multiple times, which is not supposed to be possible at all. We have been investigating this issue with the developers since then, but the issue, which affects only a small minority of users, proves difficult to replicate on the developers' machines. On the 30th of August, this issue concerned 51 players. We are waiting for the developers to solve this issue. We decided we would not use this data for any analysis. [written 21/09]
05.09.2018: We completed our first analysis of the app's data, concerning the TREES project. We found that the proper conditions to carry on the project, split the two halves into sub-groups, and provoke the growth of distinct dialects, were not met. The project has been aborted. [written 21/09]
15.09.2018: We rolled out version 1.10 of the app for iOS and Google Play. It includes Portuguese (Brazil), Japanese, and Russian language versions of the game.
09.04.2019: We decided last week to retire the app slightly earlier than planned (the planned date was May 15th). The reasons are (a) traffic on the app has slowed down a lot, we still have several visits per day but the trend is downwards (b) the projects that required a one-year data collection period were refashioned to accommodate other scales (based on the progress of individual pairs not the group as a whole, like FRIENDS) or abandoned (like TREES). The website colorgame.net does not feature the links to the app stores since last week, and on Thursday the app will be discontinued.
10.04.2019: We consider that the data collection period is over since today 1 PM. We will be playing with the app with a group of students this afternoon but this data won't be kept.

5. Descriptive and exploratory analyses

The accompanying R scripts for this section can be found on this page: https://osf.io/rywsd/.

This document records simple exploratory analyses on players' performance and frequentation, which were not part of preregistered studies. They are presented in the form of a research diary. No preregistration was made, but analyses were recorded as exhaustively as possible. The analyses were carried out on preliminary versions of the dataset, not on the complete and cleaned-up Color Game dataset. The code we provide should nevertheless be sufficient to replicate our analyses approximately with the final version of the dataset.

22/06/2018. **Players' performance in the Color Game through time.**

We looked at player performance to see whether it improved with time. It does (**Fig. 8**).

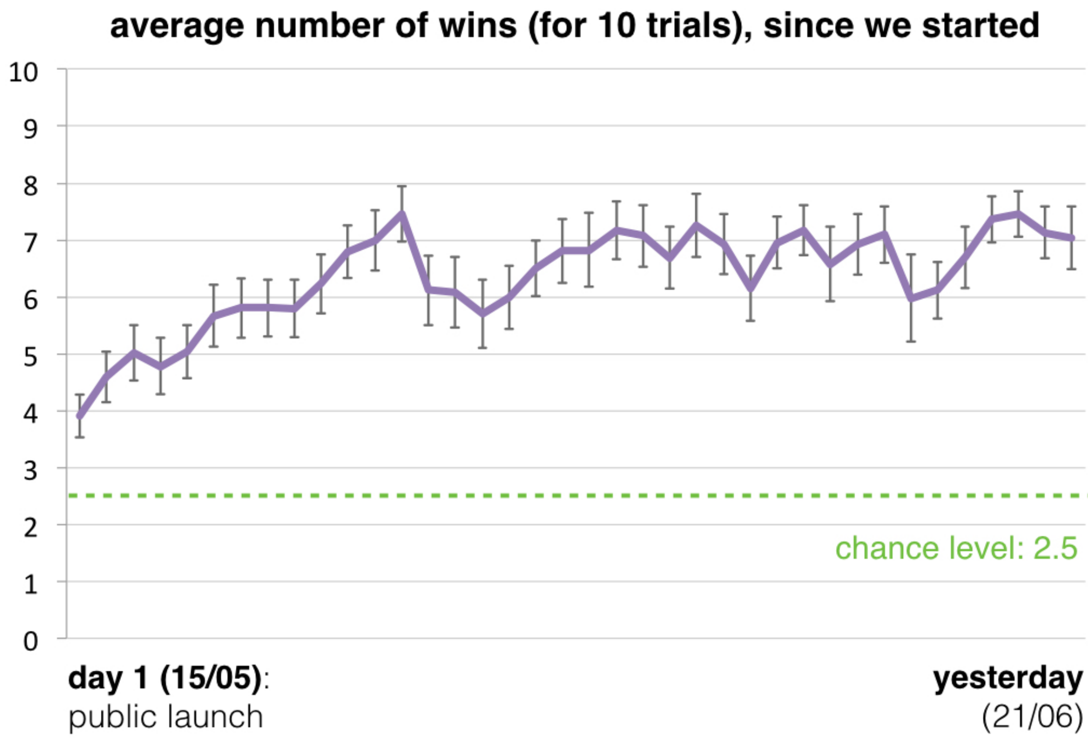


Fig. 8. The average number of correct trials for a set of 10 went from 4 to 7 in the last five weeks. The error bars represent simple variance (not standard deviations). n = 17832 sets of 10 trials.

In the next analysis, it could be interesting to tease apart three possible mechanisms for this progress.

A. Players become better decoders. In this scenario, players *simply* become better at solving any puzzle at all. They become better Receivers, but not necessarily better Senders.
B. Messages become more intuitive. In this scenario, players *simply* become better at producing solvable puzzles, whoever the puzzles are destined for. They become better Senders, not necessarily better Receivers.

C. Players become better at mastering conventions. In this scenario, players become better at using the conventions that they developed for the game; since these are purely conventional, however, this skill only works with players who have also mastered these conventions, not with new players. Players do not become better at solving or creating puzzles successfully with anyone, only with experienced players.

A + B. Convention-independent progress. In this scenario, hypotheses A and B (but not C) are both true. Players become better at solving (any) puzzle, and they also become better at creating puzzles that can be successfully solved by any player.

A + B + C. All of the above. In this scenario, all the above hypotheses are true. Players become better at solving (any) puzzle, and they also become better at creating puzzles that can be successfully solved by any player. In addition, they are even more successful when playing with another experienced player.

20/07/2018. Pairs of players' performance as a function of experience and joint experience.

Code:
- *2018-07-20-pairXP.R* for the data; a few days later we did an updated version, 2018-07-25-pairXP.R.
- *Figure-2018-07-20.R* for the figure.

A pair's performance in the game appears to be driven by the players' experience as a pair, more than the total experience of each player (playing separately), suggesting a crucial role for pair-specific conventions.

We wanted to check that the players' progress in the Color Game was due to the establishment of shared conventions—that players didn't simply improve by playing a lot, independently of establishing conventions with other players. To make sure of this, we considered two variables that could affect the performance of a pair of players. One is the pair's experience as a pair: the number of trials that the two players composing the pair played together. The other is the pair's players' total experience: the number of trials played by either player of the pair, together or separately.

To avoid the confounding effects introduced by the fact that players gain new keyboard symbols with experience, we only studied all the pairs of players where one of the players had, at some point, reached the level at which all 35 symbols are available. Looking at the whole dataset from 26/04 to 22/06, this means 4074 pairs, made up of 1412 distinct players.

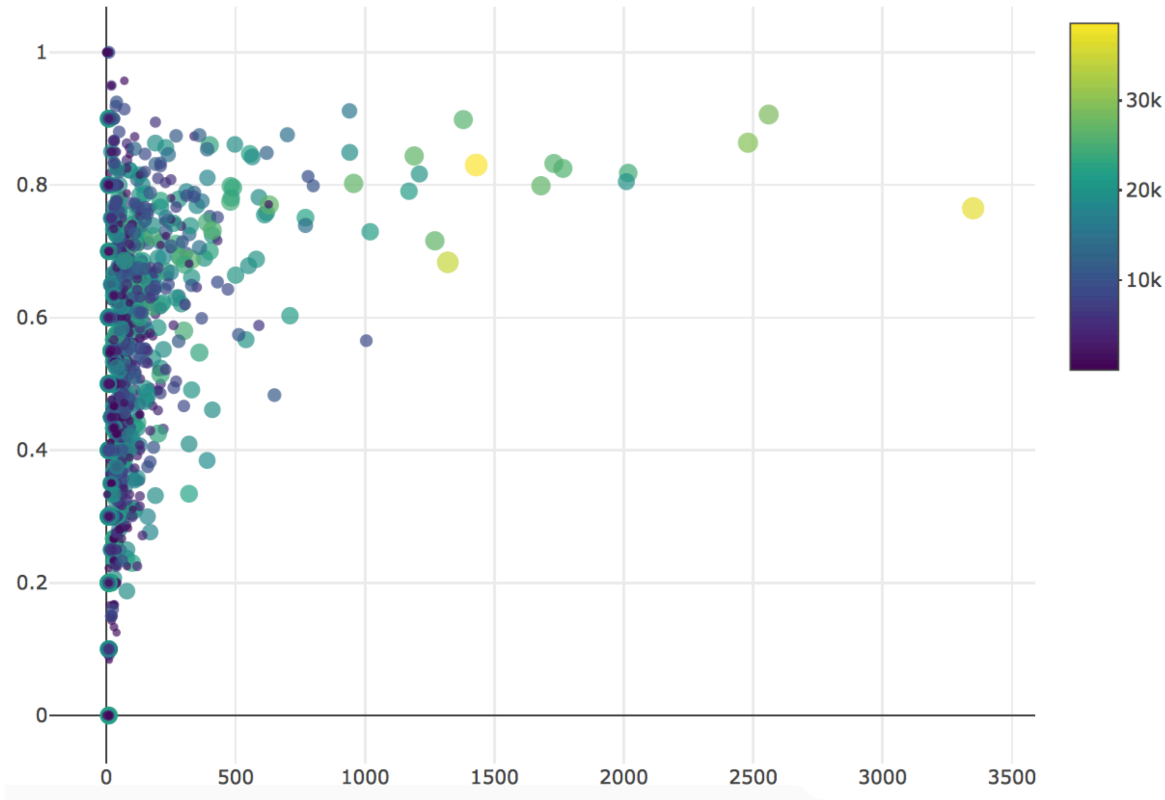


Fig. 9. Player pairs' performances as a function of their past performance playing together as a pair. Each dot stands for a pair of players (n = 4074 pairs, from 1412 players). Y-axis: the pair's chance of scoring a hit (0 to 1, chance level = 0.25). X-axis: The number of trials played by the pair as a pair. The colour and size of each dot represents the pair's general experience: The number of trials played by either player, separately or jointly.

Fig. 9 shows the result. The full figure was made with *plotly* and has scroll-over features not accessible here (but see R code). It clearly shows that joint experience is correlated with a performance increase. This could be due to a tendency for players to keep playing with the partners they achieved high scores with. But this seems unlikely: the pairs with very high joint experience perform at extremely high levels (80% success rates being common) which pairs with less joint experience seldom reach.

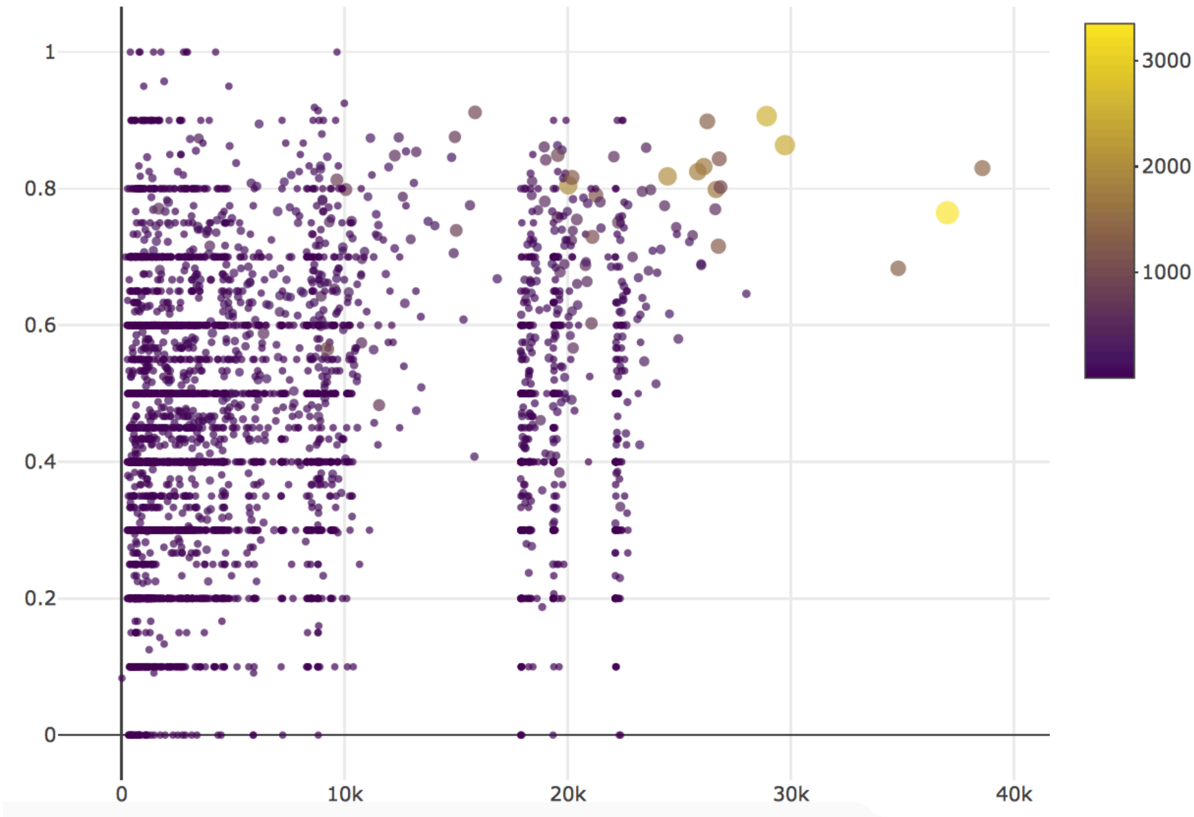


Fig. 10. Player pairs' performances as a function of their past performance playing together as a pair. Each dot stands for a pair of players (n = 4074 pairs, from 1412 players). Y-axis: the pair's chance of scoring a hit (0 to 1, chance level = 0.25). X-axis: The number of trials played by the pair's players, separately or together. The colour and size of each dot represents the pair's joint experience: the number of trials played by either player, separately or jointly.

The experience of players outside the pairs, in contrast, does not seem to matter as much, even though the variance for this variable is much greater. This is shown in **Fig. 10**, which plots performance against a pair of players' general experience, gained together or separately. Most performance scores cluster around round values (0.1, 0.2, etc.), because many pairs only play 10 or 20 trials together. The odd shape of the data in the 10K – 20K trials range is probably due to the fact that a few highly experienced players engaged a lot of players with very little experience.

This was also a good occasion to see what share of the game's activity could be attributed to what number of players, and identify possible problems with the data.

14/08/2018. Another look at pairs' performances.

Code:
- *PairExperience.R*

We wrote up a much more elegant version of the code used for the preceding entry. We did a re-run of this code on the updated dataset. Results in **Fig. 11–13**.

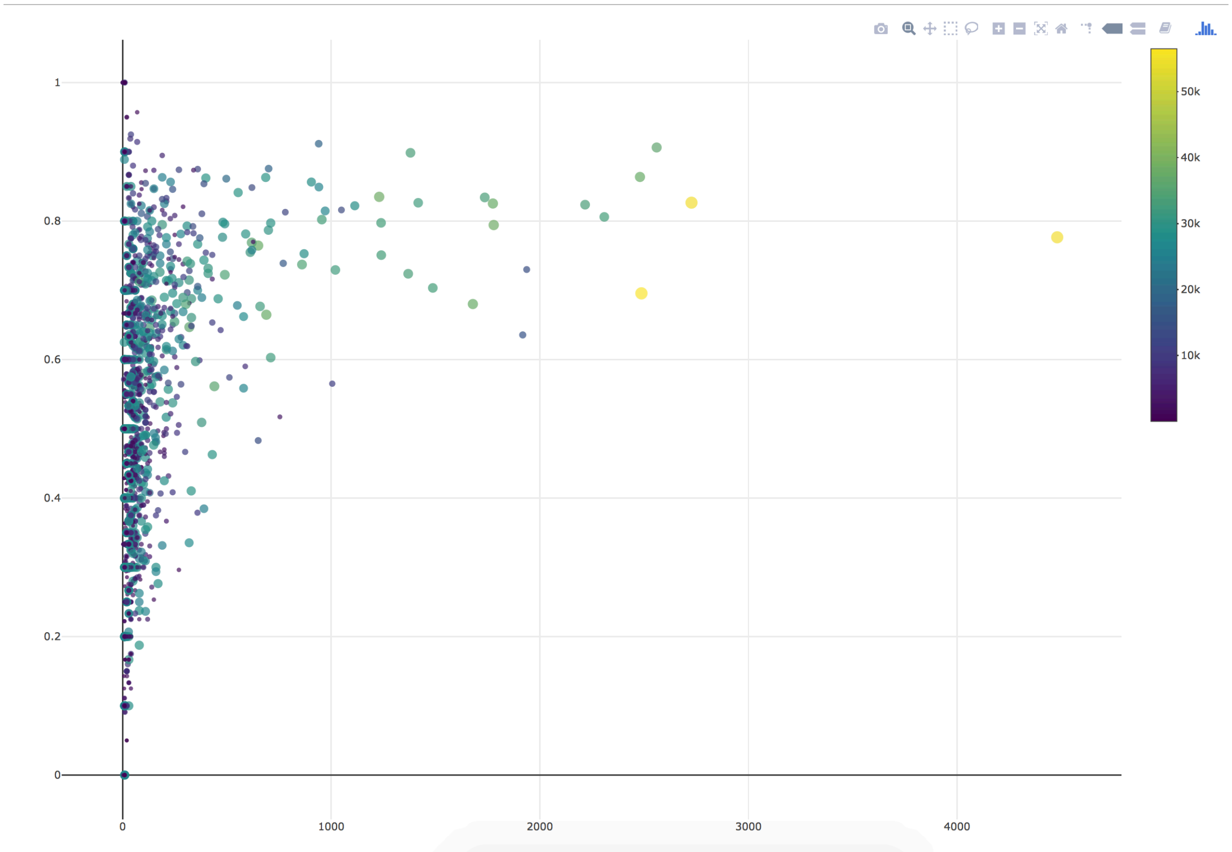


Fig. 11. Player pairs' performances as a function of their past performance playing together as a pair. Each dot stands for a pair of players (n = 5506 pairs, from 1528 players). Y-axis: the pair's chance of scoring a hit (0 to 1, chance level = 0.25). X-axis: The number of trials played by the pair as a pair. The colour and size of each dot represents the pair's general experience: the number of trials played by either player, separately or jointly.

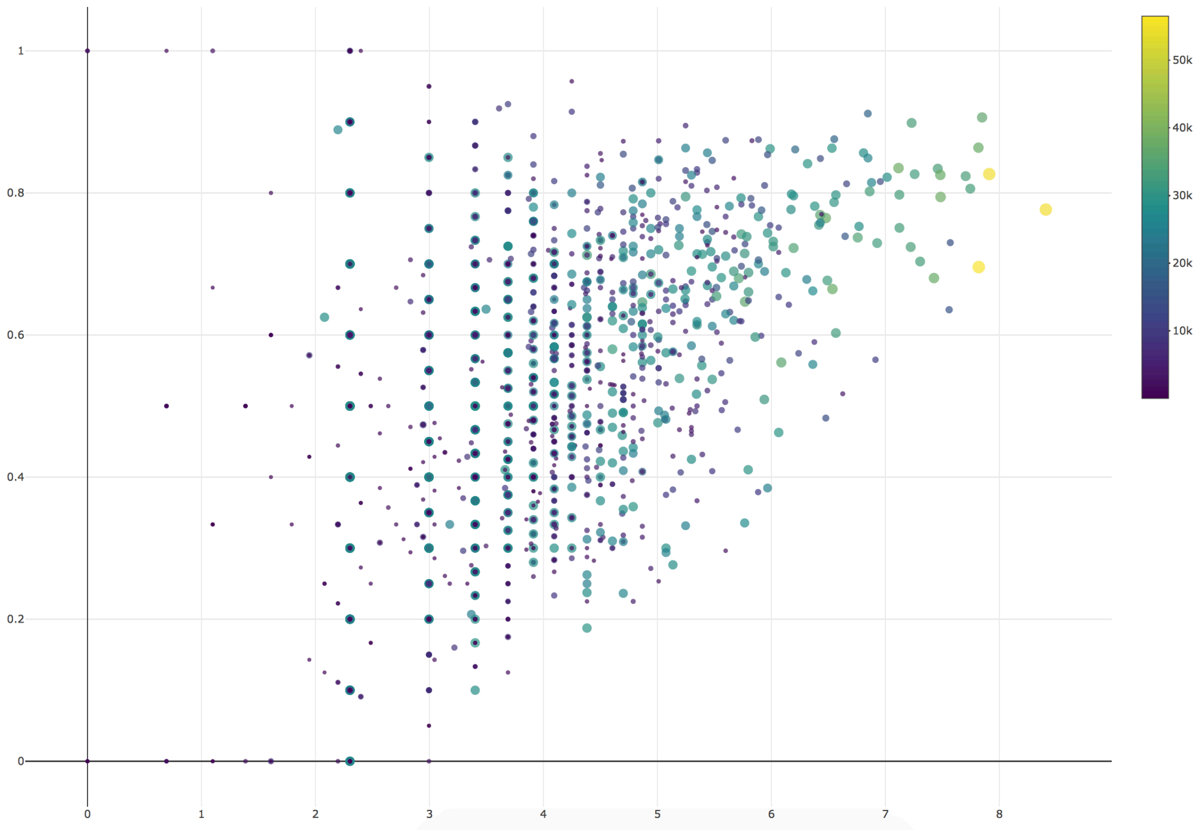


Fig. 12. The previous figure, with the joint experience log-transformed.

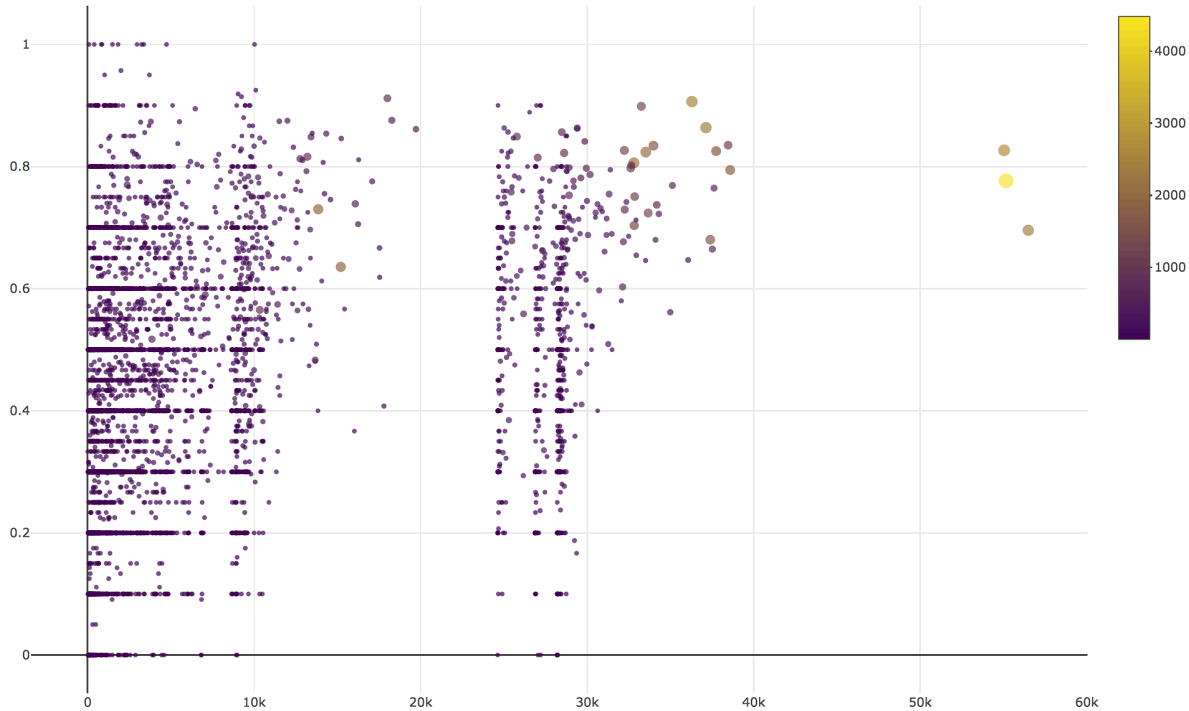


Fig. 13. Player pairs' performances as a function of their past performance playing together as a pair. Each dot stands for a pair of players (n = 5506 pairs, from 1528 players). Y-axis: the pair's chance of scoring a hit (0 to 1, chance level = 0.25). X-axis: The number of trials played by the pair's players, separately or together. The colour and size of each dot represents the pair's joint experience: The number of trials played by either player, separately or jointly.

06/09/2018. Traffic and frequentation.

Code:
- *CGFrequentation-2018-08-15.R*, last updated 2018-08-15.
- *CleanUp-2018-09-06.R.*

We mapped the frequentation of the app: the number of players, number of trials, when they play and how often, how they perform, and their favourite time of day for playing. **Fig. 14, 15, and 17** cover the time span from May 15th to September 5th, 2018 on the x-axis. **Fig. 16** shows player affluence by time of day.

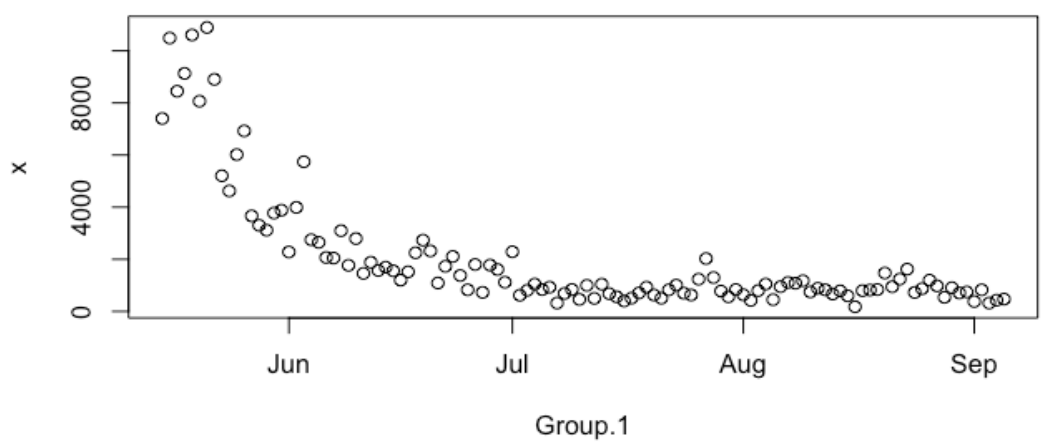


Fig. 14. Number of trials per day.

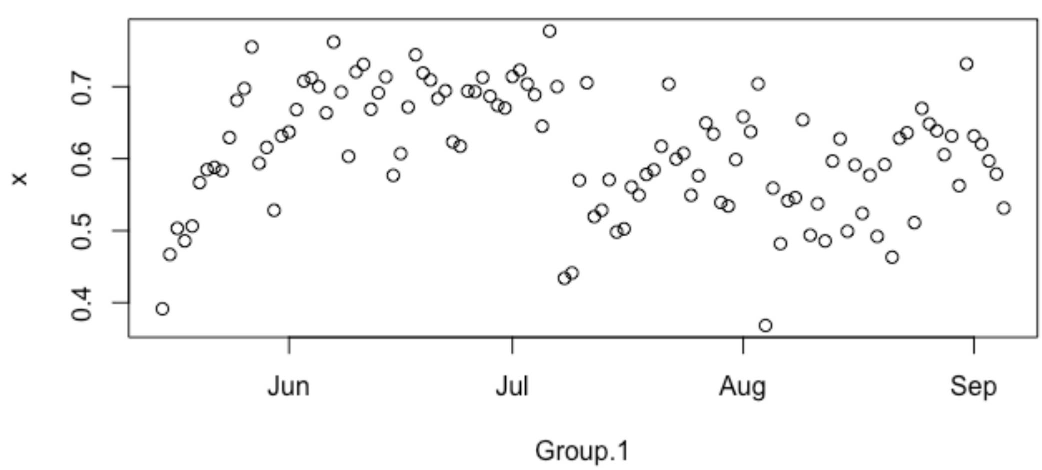


Fig. 15. **Average performance rates per day.** The performance increase seen on 22/06/2018 (see that entry) did not continue.

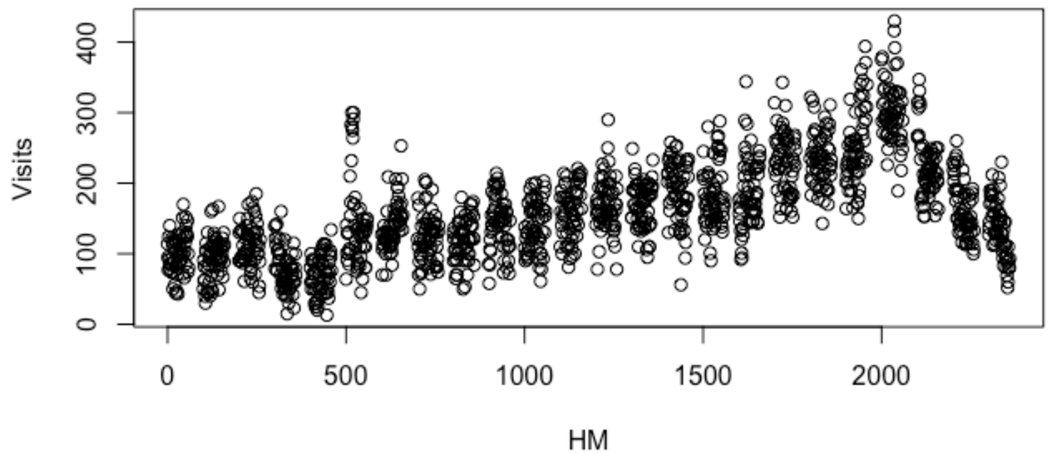


Fig. 16. Number of trials played on the app (y-axis) for each minute of the day (x-axis). Time is given as Universal Time. Central European Time is two hours later (20:00 = 22:00).

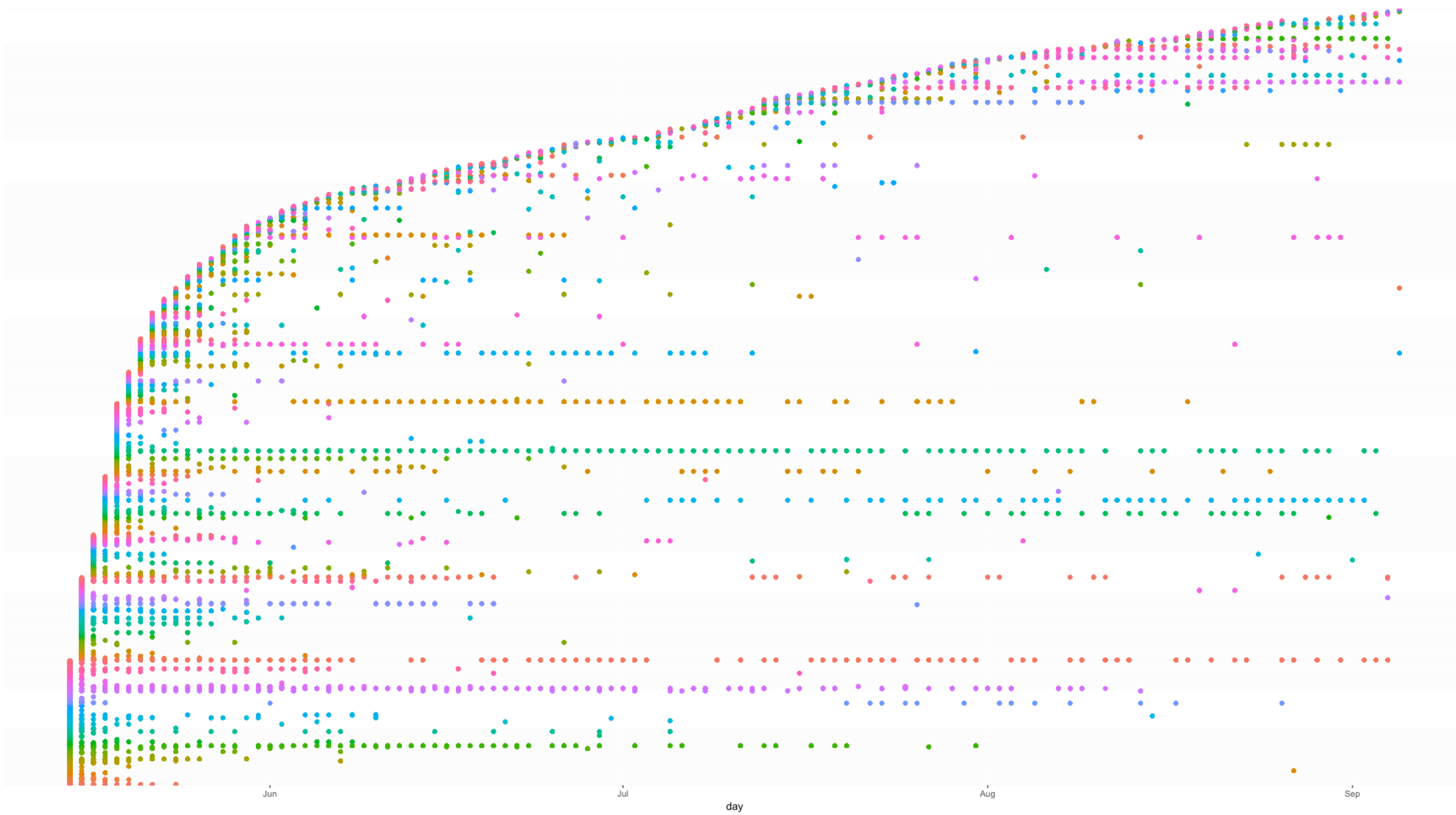


Fig. 17. Players visits to the Color Game. Each row represents one player, with the earliest visitors shown at the bottom. Each dot represents one visit on one day, by one player.

02/10/2018. Traffic by hour.

Code:
- *CGFrequentation-2018-10-02.R.*

We plotted the peak traffic times of the day with a circular plot (**Fig. 18**).

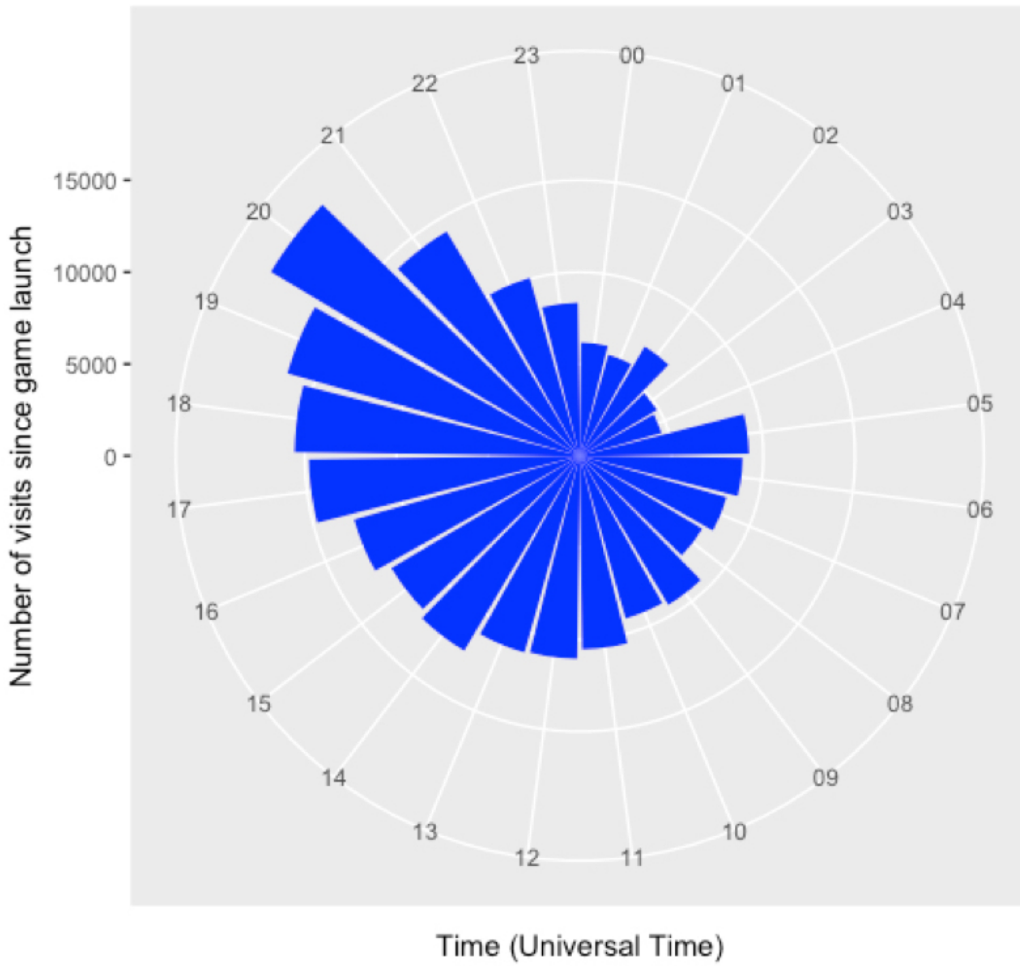
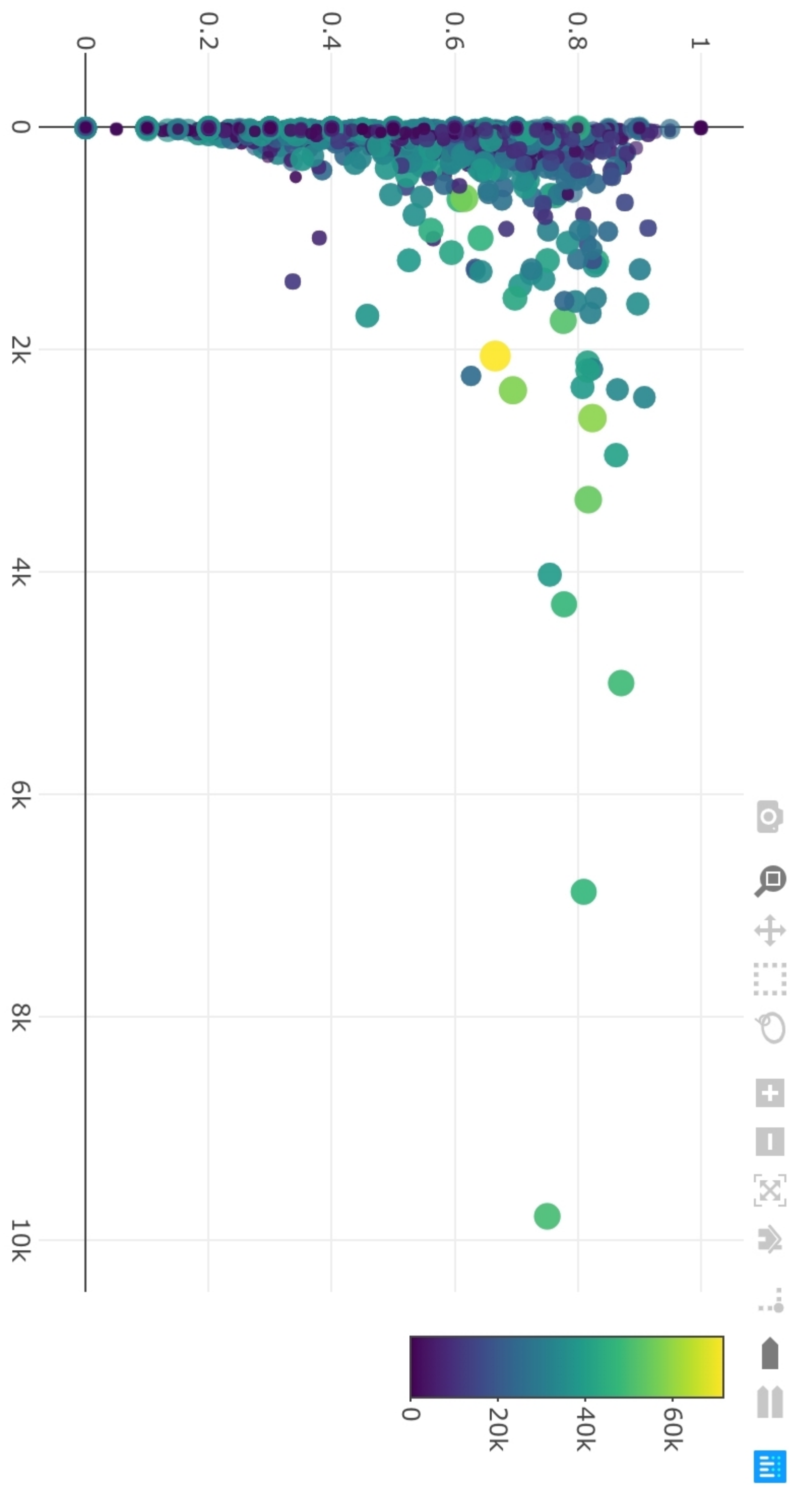


Fig. 18. Players affluence by time of day, on a circular plot.

05/09/2019. Replication of Fig. 11, with the final data.



6. Acknowledgements

The Color Game app was designed by the Minds and Traditions Research Group within the Department for Linguistic and Cultural Evolution of the Max Planck Institute for the Science of Human History in Jena, in collaboration with Etter studios (Zurich).

The basic principle of the color game was an original idea of Olivier Morin and Tiffany Morisseau in 2015. The game was then developed and investigated experimentally by Olivier Morin, Thomas Müller, James Winters, Lidiia Romanova at the Minds and Traditions Research Group of the Max Planck Institute for the Science of Human History.

The blueprint for the app was conceived by Olivier Morin, James Winters, Thomas Müller, Tiffany Morisseau, Lidiia Romanova, and Christian Etter. We are grateful to Helene Kreysa and Simon Greenhill for their helpful comments. The blueprint was implemented by Etter studios in Zurich under the supervision of Christian Etter.

The technical supervision for server maintenance and data storage in Jena was assured by Thomas Baumann and his team. The website colorgame.net is administered by Syahrul Anwar at the institute.

The app's symbols were designed by Thomas Müller, Lisa Jeschke, Liliana Lovallo, Barbara Pavlek, Piers Kelly, James Winters, Olivier Morin, Lidiia Romanova, Olena Tykhostup. The game's graphic design is the joint work of Michelle O'Reilly, Olivier Morin and James Winters, in collaboration with Etter studios.

Translations were the work of Thomas Müller, Olivier Morin, Lisa Jeschke, Charlélie Goldschmidt, Ezequiel Koile, and Tao Li — with special thanks to Messrs. Koile and Li.

We are grateful for the help of our Beta-testers. In no particular order: Teresa Blasco, Alina Shron, Pierrick Bourrat, Réka Finta, Abel Gerschenfeld, Charlélie Goldschmidt, Liem Binh Luong Nguyen, Olivier Mascaro, Pierre-Yves Le Coq, Barbara Pomiechowska, András Szanto, Radu Umbres, Laëtitia de Villepin, Rosie Ivády, Chris Winters, Katie Finch, Matt Spike, Stefan Becker, Suzan Huijgen, Maria Brackin, Seán Roberts, George Dunford, Claudia & Stefan Geßlein, Thomas Borchert, Sarina Parschick, Mathias Löbl, Anneke Hamann, Lisa Kerscher, Jana Koplanski, Julia Belger, Judith Lidzba, Elisabeth Wienß, Susanna Sabin, Ezequiel Koile, Sarah Wand, George Taylor, Ian Gillespy, Raul Azevedo, Camille Kerdanet, Cécile Eymond, Johan Morisseau, Maximilien Quesnel, Jérôme Redouté, Laurent Jacquod, Ira Noveck, Katell Lavéant, Katrina & Barbara Pavlek.

The game's promotion is assured by the Grafikerorg agency (Jena). We are grateful to Aditya Kumar Lankapalli, Eirini Skourtanioti, Hiba Babiker, Cosimo Posth, and Xintian Li, who acted in the advert.

Inside the Max Planck Institute for the Science of Human History, we are especially grateful for the help of Lidiia Romanova, Lisa Jeschke, Anne Gibson, Petra Mader, Christian Nagel & Barbara Pavlek.

The work leading to this paper has received funding from the "Frontiers in Cognition" EUR grant, ANR-17-EURE-0017 EUR.

7. Creative Commons Licence

8. Ethics and data privacy

The following information was displayed on the colorgame.net website, as part of a downloadable pdf document that also contained the official ethical approval for the project.

"The Color Game app stores smartphone identifiers securely and confidentially. It also records, in fully anonymised format, the players' actions within the game. This fully anonymised data will be shared with the scientific community, but no personal data will be. Participants are free to opt out of our research and ask for their data to be erased. This study has been approved by the ethical committee attached to the president of the Max Planck Society.

The app identifies its players' phones (via UDID on iOS, UUID on Android) to make sure that players do not have to restart the game at the basic level every time they open the app. The app also checks its players' IP addresses (but does not record it). These data, along with all the data used by the Color Game app, are exclusively stored on the secure servers of the Max Planck Society. The app does **not** require or record the following information: names, email addresses, geolocalisation tags.

Inside the game, players identify their contacts by randomly generated aliases that change for every player pair (so that A is known to be by a certain name, but known to C by a different name), making it impossible for contacts to identify one another outside the game. In addition, players are identified on the scoreboard by a pseudonym that is randomly generated by the app (although players are free to reject pseudonyms that they dislike).

The researchers of the Max Planck Society access a fully anonymised subset of the data generated by the app, which they store and exploit for research purposes, and may share with other researchers. It is a record of the players' moves in the game in anonymised format. For instance, the colours picked by Receivers and the symbols sent by Senders are recorded for every puzzle. The time at which puzzles are played is also recorded. The IP addresses, UDID/UUID, aliases, and pseudonyms will **not** be part of that dataset kept for research purposes.

Players are asked to fill in a consent form allowing their player data, collected through the app and entirely anonymised, to be used by researchers for scientific purposes.
The data collected during the period of time where the app will be in use is to be made publicly available to other researchers in the future, after an embargo period. Any player wishing to have their data removed from the dataset may ask us to do so (at colorgame@shh.mpg.de) and their request will be immediately fulfilled. (Since the app's anonymised IDs do not allow us to identify individual players, the players will need to provide us with a way to identify them: for instance, the exact date and time of their first connection, or the exact date and time of their last five puzzles played.)

This study has been approved by the ethical committee attached to the president of the Max Planck Society (advice n° 2017_05). The advice is attached to this document, followed by our official request for authorisation."

References

Luo, M. R., G. Cui, and B. Rigg
2001 The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000. Color Research & Application 26(5): 340–350.

Morin, Olivier, James Winters, Thomas F. Müller, et al.
2018 What Smartphone Apps May Contribute to Language Evolution Research. Journal of Language Evolution 3(2): 91–93.

Müller, Thomas F., James Winters, and Olivier Morin
2019 The Influence of Shared Visual Context on the Successful Emergence of Conventions in a Referential Communication Task. Cognitive Science 43(9): e12783.

Yule, George
2013 Referential Communication Tasks. Routledge.