Ranking-based social influence produces path-dependence

Keywords: social influence, path-dependence, Matthew effect, rich-get-richer, ranking effects

Introduction People use social cues to inform their choices. It is well known that in binary or multi-alternative choice, information cues about the various options' popularity can increase the probability that an agent chooses already popular options (*1*), possibly making an inferior option more likely to be selected than a superior one. This observation has generally been thought to imply that social influence processes are unpredictable, in the sense that randomness early on can substantially affect the long-term outcome, a property also known as path-dependence. In particular, lock-ins to states where an inferior option becomes the most popular one are possible (*2*).

However, the belief that most social influence processes are path-dependent has been recently challenged by van de Rijt (*3*). Re-analyzing a variety of datasets, the author found that several cue-based social influence processes that were previously thought to be path-dependent were in fact self-correcting; a superior option would eventually always become more popular, even if the inferior one had a significant initial advantage. This led the author to suggest that social influence in most realistic settings cannot possibly lead to cumulative advantage. By contrast, in a recent experiment on political opinion formation conducted by Macy and colleagues (*4*), all the signatures of path-dependence and lock-ins were found, with no evidence of self-correction. Path-dependence was also implicit in an experiment by Frey & van de Rijt (*5*), where it was found that the majority beliefs of groups of people were more likely to be wrong in the presence vs absence of social influence. These incongruent findings raise the following question: how broadly should we expect lock-ins to occur in the real world and what are some properties of the system that make lock-ins more likely?

Theoretical results We consider a setting previously studied by Arthur (*2*) and that is suitable for modeling the above experimental settings: a large number of individuals are presented sequentially with a choice between two options, A and B. The probability p_A that a subject chooses option A is a function of the current popularity of A, defined as the proportion of previous agents that have chosen this option. That is, there is some function f such that $p_A = f(x_A)$, where $x_A = \frac{n_A}{n_A+n_B}$ is the current popularity of A (fig. 1). We call f the *influence curve*. We assume that f is increasing and that, without loss of generality, $f(0.5) \leq 0.5$, implying that option A is inherently (equally or) less appealing than option B.

Under some mild technical conditions on f, the popularity of A converges as the number of agents grows large, but the limit $\bar{x}$ is ex ante random. We call the system *self-correcting* if $\mathbb{P}(\bar{x} > 0.5) = 0$, i.e. if the probability that A (the inferior option) will be chosen by a majority of agents in the long-term is 0. Otherwise we say that the system is *path-dependent*. The following theorem[1] says that we can find the possible long-term limits for x_A and whether the system is self-correcting or path-dependent just by looking at the influence curve (see fig. 2).

Theorem. 1. *For any* $x_0 \in (0,1)$, $\mathbb{P}(\bar{x} = x_0) > 0$ *if and only if the graph of* f *downcrosses the line* $y = x$ *at* $x = x_0$, *i.e.* $f(x) > x$ *in* $(x_0 - \varepsilon, x_0)$ *and* $f(x) < x$ *in* $(x_0, x_0 + \varepsilon)$.

2. *The system is path-dependent if and only if the graph of* f *crosses* $y = x$ *at some* $x > 0.5$.

[1] Part 1 of the theorem follows from results in (*2*). Part 2 appears implicitly in (*3*).

Next we assume a discontinuity for f at $x = 0.5$ and write $f(x) = g(x) + \frac{R}{2} \cdot u(x)$, where g is a continuous increasing function, $u(x)$ is a step function that takes the value 1 or -1 depending on whether $x > 0.5$ or $x < 0.5$, and R is a constant that we call the *ranking effect* (fig. 3). We also denote by d the difference in inherent appeal of the two options, defined as the difference in choice probability of B vs A when they are equally popular, i.e. $d = (1 - g(0.5)) - g(0.5)$.

Theorem. *If $R > d$, then the system is path-dependent.*

This theorem provides an alternative, sufficient condition for path-dependence and lock-ins. It says that lock-ins to the inferior option are possible in systems in which the ranking effect is large enough to overcome the quality difference of the two options. We thus predict that lock-ins are likely if there is a discontinuity in the influence curve at $x = 0.5$, i.e. if people react to which option is more popular independently of the size of the difference in popularity.

Datasets We analyze data from three binary choice experiments under social influence (*3*–*5*) and estimate the influence curves and ranking effects for each of the experimental items (questions). In (*3*), subjects were asked 7 questions on matters of taste and were shown the number of people before them who had chosen each of the two possible answers. One trial of the experiment involved 530 participants and another 3500 participants; the latter gave an artificial initial advantage to the inferior option (determined in a control experiment) of $n_A \approx 110$ to $n_B \approx 10$. In (*4*) participants in the US that self-identified as democrats or republicans were asked whether they supported 20 statements. For each statement they were informed whether it had so far a larger support by democrats or republicans, but not by how much more. There were 10 trials of approx. 220 participants each (+ control condition). In (*5*) subjects were asked 30 questions that had a ground truth and were incentivized to answer correctly. They were informed of the number of previous participants that had given each of the two possible answers. There were 30 trials (+ control), half with 100 participants and half with 15.

Empirical results We find that in (*3*) the influence curves never cross the diagonal for $x > 0.5$ and, as predicted, no lock-ins are observed (fig. 4). In (*4*), by contrast, the diagonal is crossed on the right half of the graph for 19 out of 20 questions and lock-ins are observed in the same 19 questions (fig. 5). Finally, in (*5*) the diagonal is crossed for $x > 0.5$ in 8 out of 30 questions (fig. 6). Lock-ins are observed in 7 out of 8 cases where they are predicted and in 5 out of 22 cases where it is not predicted.

In figs. 4 to 6 strong ranking effects are evident in many of the questions. Figure 7 shows the questions from all datasets in the R–d space. The ranking effect R is estimated as the increase in the influence curve between the two central bins (0.4-0.5 and 0.5-0.6), while the quality difference d is estimated from the control data (no social influence). Although the condition $R > d$ was found theoretically to be merely sufficient for lock-ins, in these datasets it turns out to be close to necessary as well, correctly classifying 51 out of 57 items. Thus, its accuracy as a predictor is in par with the influence curves, despite requiring much less data to be estimated.

References

1. M. J. Salganik, P. S. Dodds, D. J. Watts, *Science* **311**, 854–856 (2006).

2. W. B. Arthur, *The Economic Journal* **99**, 116–131 (1989).

3. A. van de Rijt, *American Journal of Sociology* **124**, 1468–1495 (2019).

4. M. Macy, S. Deri, A. Ruch, N. Tong, *Science Advances* **5**, eaax0754 (2019).

5. V. Frey, A. van de Rijt, *Management Science* **67**, 4273–4286 (2021).

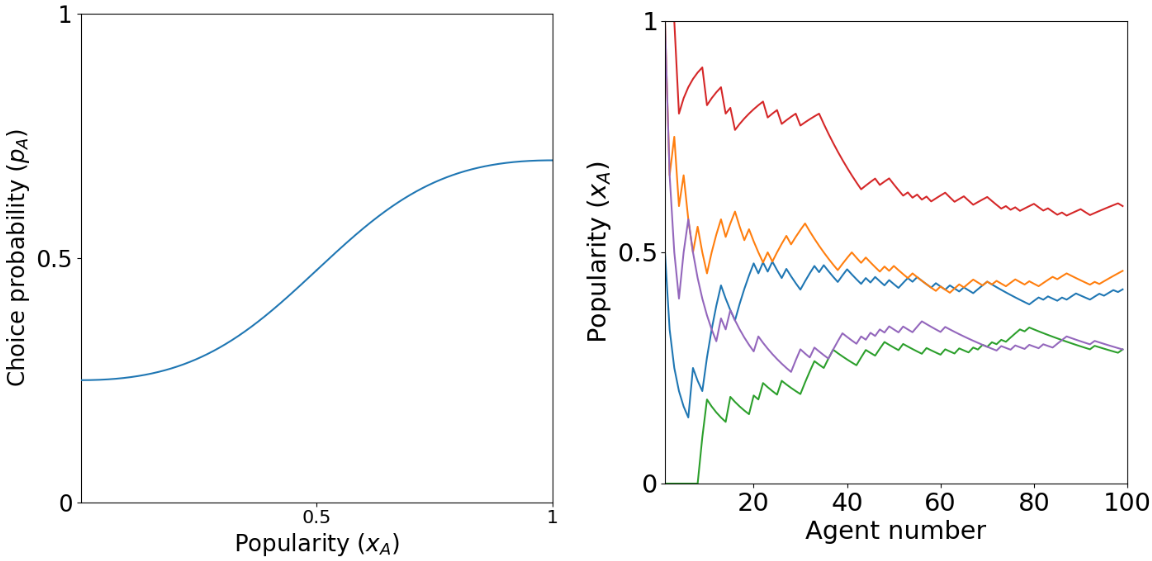

Figure 1: **Left:** In binary choice experiments, the probability that an agent chooses item A can be modeled as a function of the current popularity of A, hence it can be represented by an influence curve. **Right:** Simulations of 100 agents choosing sequentially, based on the influence curve shown on the left. Initially the trajectories fluctuate substantially, but they eventually stabilize.

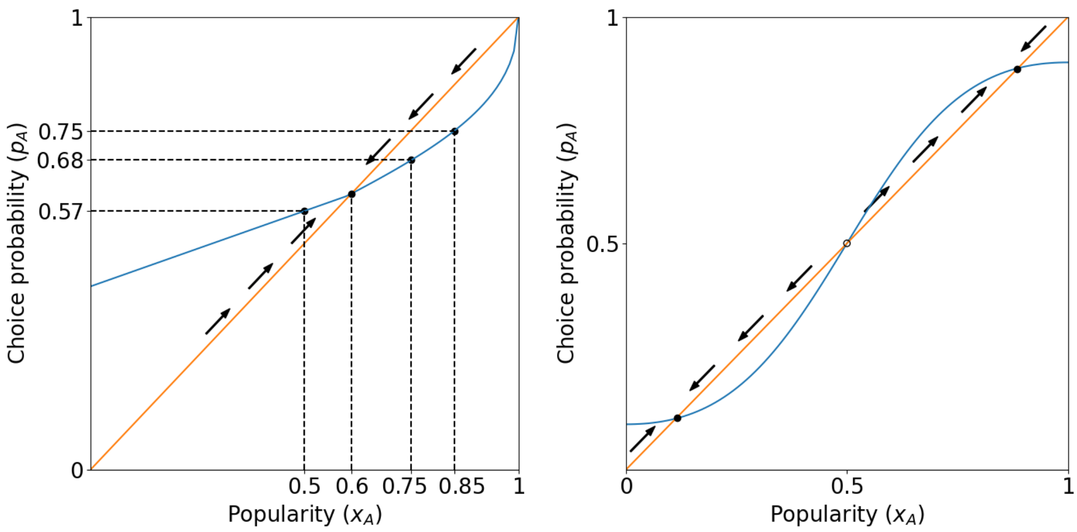

Figure 2: Illustration of Theorem 1. **Left:** In this influence curve, if the popularity of item A is $x_A = 0.5$, then it is selected with probability $p_A = 0.57 > x_A$, thus driving its popularity upwards (indicated with an arrow pointing upwards along the diagonal). If instead A's popularity is $x_A = 0.85$, it is selected with probability $p_A = 0.75 < x_A$, driving its popularity downwards. As the popularity decreases from 0.85 to 0.75, the choice probability also goes down, driving the popularity further down and so on. At the point where the influence curve intersects the diagonal ($p_A = x_A = 0.6$), the choice probability is equal to the popularity, indicating an equilibrium. **Right:** The influence curve can intersect the diagonal at multiple points, implying multiple possible equilibria. At the points where the graph crosses the diagonal *downwards* (passing from above the diagonal on the left side to below the diagonal on the right side – filled circles), the equilibria are stable, meaning that deviations from those will be eventually corrected (arrows showing the direction that popularity moves are converging). When it crosses upwards (open circle), there is an unstable equilibrium (popularity from nearby points moves away from that point). If the popularity starts near the unstable equilibrium, initial randomness can lead it to converge to either of the two stable equilibria.

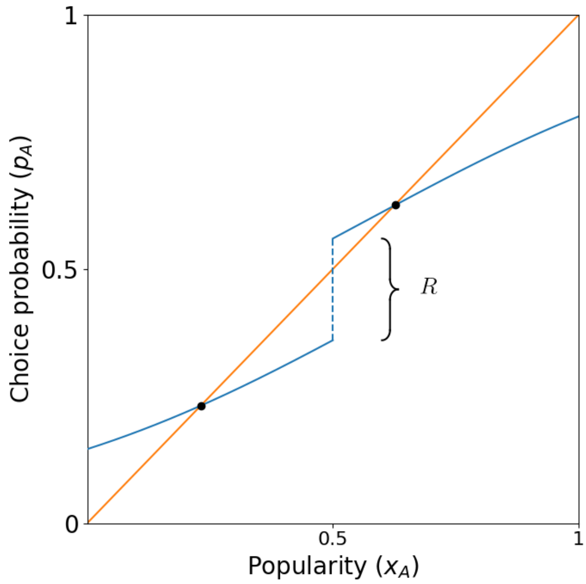

Figure 3: An influence curve with a discontinuity at $x = 0.5$ can be decomposed into a continuous part and a ranking part, $f(x) = g(x) + \frac{R}{2} \cdot u(x)$, where $u(x) = \pm 1$ if $x \gtrless 0.5$. The ranking effect R equals the jump at $x = 0.5$.

van de Rijt (2019)

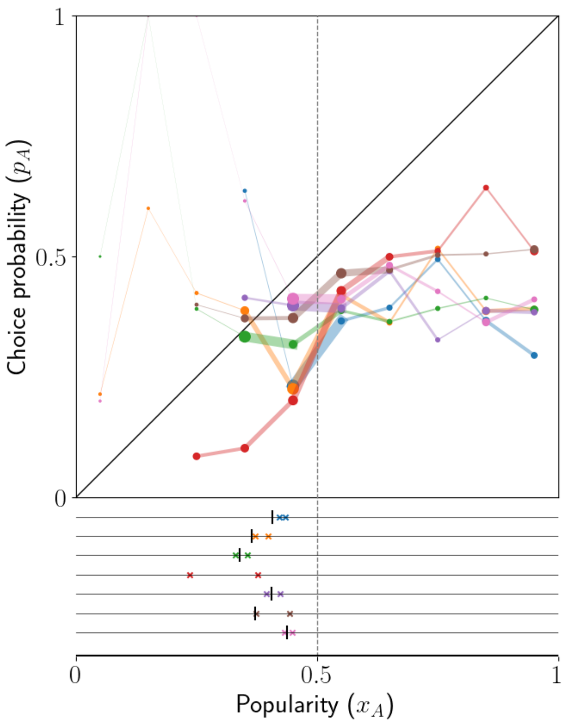

Figure 4: Influence curve estimation for the experiment of van de Rijt (*3*). For each question, answers from all participants in all trials are pooled and split into 10 bins of equal size, based on the popularity x_A of the inferior option at the time (in the participant's world/trial). Among the answers that fall in each bin, we find the proportion of times that option A was chosen. The size of each circle is proportional to the number of data points in the interval. Similarly, the width of each line is proportional to the minimum number of data points in the two intervals that it connects. Given that none of the influence curves crosses the line $y = x$ for $x > 0.5$, the theory predicts that lock-ins are impossible. The horizontal lines at the bottom (one per question) indicate the end-of-trial popularities ($\bar{x}$ – colored crosses) of option A in the two trials of the experiment. In all cases $\bar{x} < 0.5$. The points where the influence curves downcross the diagonal are indicated with vertical bars. These points are the predicted long-term proportions of A and they are agree reasonably with the observed end-of-trial proportions. Note that in one of two trials for each question option A was given a large artificial advantage in popularity, hence the lack of a lock-in is strongly indicative of its impossibility, as noted in (*3*).

Macy et al. (2019)

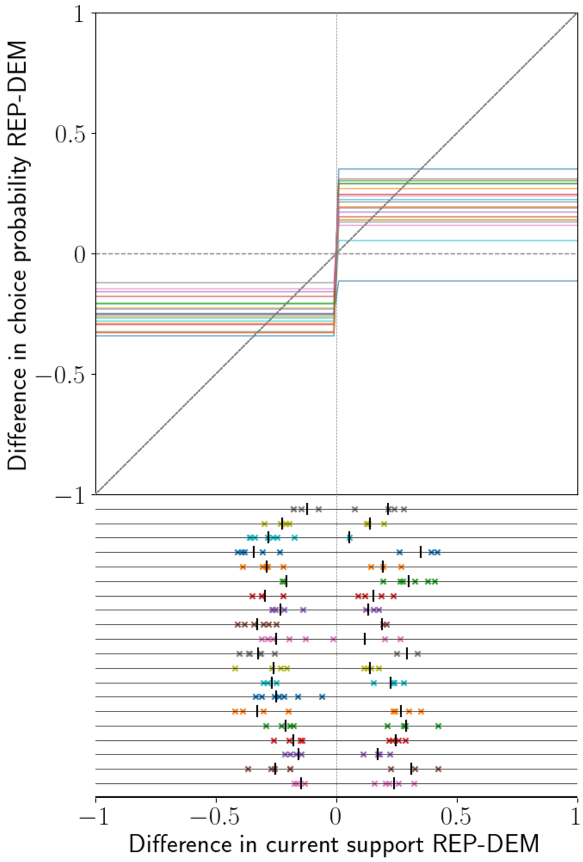

Figure 5: Influence curves in the Macy et al. experiment (*4*). In this experiment the participants were only informed whether the support for a statement so far was larger for democrats or republicans, hence the choice probability is constant no matter the size of the majority (flat lines on either side of $x = 0$). The variables of interest are now in the range $[-1, 1]$ instead of $[0, 1]$, but the theory still applies, with obvious modifications. In all but one case the lines cross the diagonal at some $x > 0$, predicting that lock-ins are possible. We find that in all but the same one case, end-of-trial majority support by either democrats or republicans was observed in at least one of the trials.

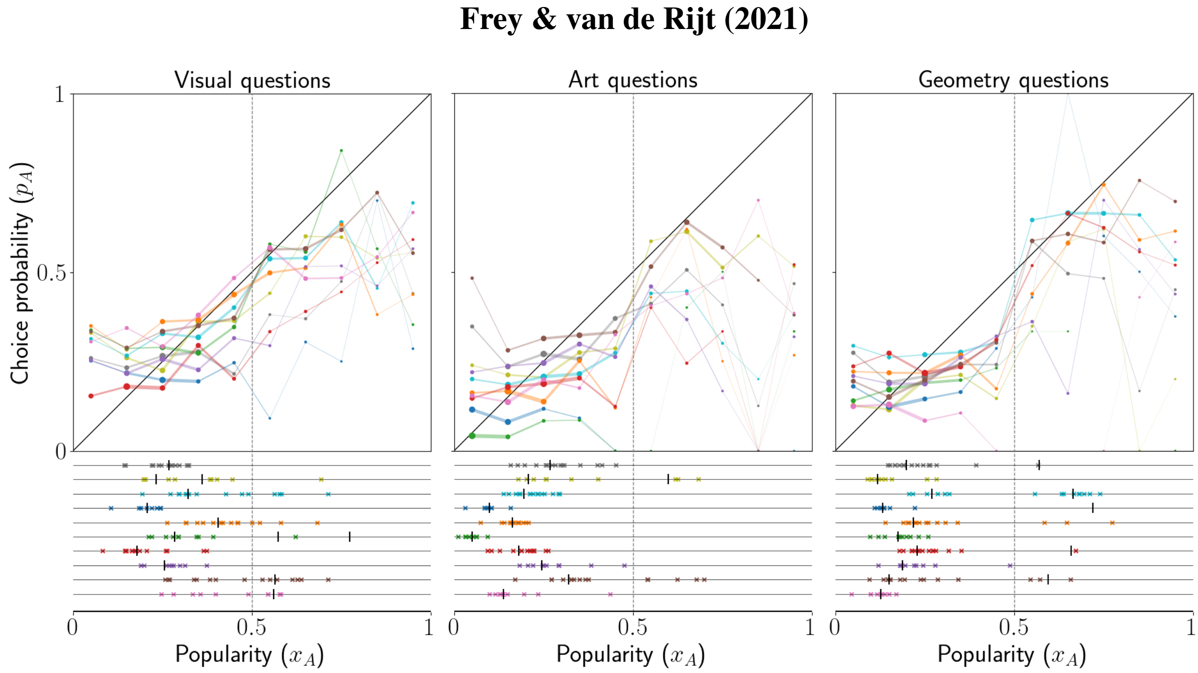

Figure 6: Similar to fig. 4, but for the experiment of Frey and van de Rijt (*5*). In this experiment one of the two possible answers for each question was correct and participants were incentivized to answer correctly. There were 10 visual judgement questions (**left**), 10 art history questions (**middle**) and 10 geometry questions (**right**). Across panels, the diagonal is crossed at some $x > 0.5$ in 8 cases and in 7 of them end-of-trial majorities of option A are observed (in at least one of the trials), which is evidence for lock-ins. There are also 5 cases where the diagonal is not crossed, but end-of-trial majorities for A are still observed (false negatives). There are prominent ranking effects in the geometry questions, and to a lesser extent in the art history questions.

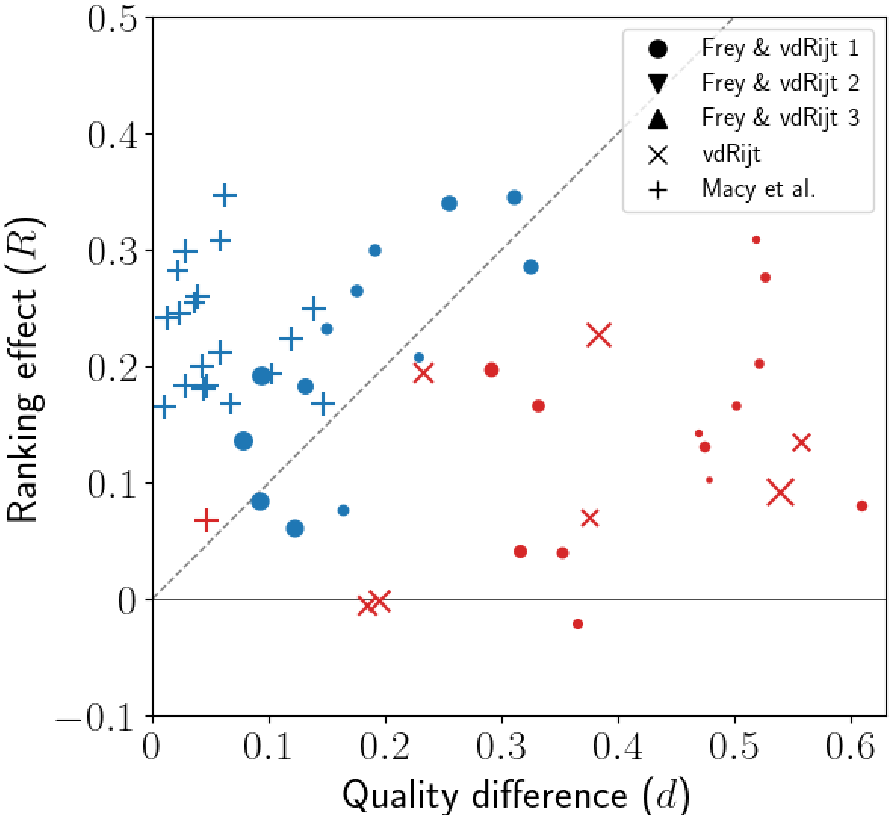

Figure 7: Summary of ranking effect vs quality difference, for all datasets. Each data point corresponds to one experimental item (question). The ranking effect R is estimated as the increase in the influence curve between the bins 0.4-0.5 and 0.5-0.6 (see fig. 4). The quality difference d is estimated as the difference in choice probability of the two options in the control experiments (no social influence). Blue points correspond to questions in which lock-ins were observed, i.e. option A had an end-of-trial majority in at least one of the trials, while red color indicates the absence of lock-ins. The theory predicts that for all points above the dashed gray line ($R = d$) lock-ins are possible, while no prediction is made for points below the dashed line. It turns out that for the datasets examined the condition $R > d$ is highly accurate even as an if-and-only-if condition for lock-ins.