

PSYCHOLOGICAL SCIENCE

Lenience breeds strictness: The generosity-erosion effect in hiring decisions

Marc-Lluís Vives^{1*}, Tania Fernandez-Navia², Jordi J. Teixidó², Miquel Serra-Burriel³

In recruitment processes, candidates are often judged one after another. This sequential procedure affects the outcome of the process. Here, we introduce the generosity-erosion effect, which states that evaluators might be harsher in their assessment of candidates after grading previous candidates generously. Generosity is defined as giving a candidate the lowest possible grade required to progress in the hiring process. Analyzing a high-stake hiring process, we find that for each candidate graded generously, the probability for subsequent candidates to pass decreased by 7.7% (experiment 1; $N = 11,281$). Testing the boundary conditions of the generosity-effect, we explore a hiring process that, in contrast to the previous process, was very selective, because candidates were more likely to fail than to pass. In this scenario, no evidence is found for the generosity-erosion effect (experiment 2; $N = 3171$). Practical implications and mechanisms underlying the generosity-erosion effect are further discussed.

INTRODUCTION

Many judgments and decisions should be made independently. These decisions, however, are rarely faced in isolation; they often come sequentially (1). One after the other, teachers grade their students or judges evaluate whether to give prisoners parole. In principle, each candidate, be they a student or a prisoner, should be evaluated solely based on their merits—or demerits—regardless of their position in the sequence. This is often not the case: People's judgments and decisions suffer from sequential effects, which substantially, and unfairly, affect ones' chances of receiving a positive or negative evaluation (2, 3). Sequence effects can thus jeopardize the fairness and efficiency of relevant deliberation processes.

Although sequential effects are a well-established phenomenon found in a plethora of contexts (4–6), doubts can be cast on whether their consequences are so critical and extended. So far, most research has examined sequential effects in individual judgments and decisions [few exceptions: (2, 3)]. However, in today's world, crucial decisions are often the result of a collective effort made by a group of experts. Collective decision making promotes normative and more accurate behavior [wisdom of the crowds (7–10)]. Here, we explore sequential effects in two hiring processes relying on the collective judgment of a committee of evaluators. In the first process, we investigate the existence of arbitrary biases in more than 10,000 assessments by 165 committees composed of five experts who evaluated candidates for a permanent public teaching position in Catalonia, Spain. In the second process, we repeat the same analysis for the evaluation of more than 3000 candidates for the position of judge in Spain, also collectively evaluated by a committee of nine experts. Establishing the existence, as well as the mechanism of sequential effects in collective judgments, is crucial to understanding the extent to which sequential effects can affect everyday life decisions.

Sequential effects have mainly been explained as the by-product of cognitive biases such as contrast effects or narrow bracketing. Simply put, the fifth candidate of the day may be less likely to obtain

a positive evaluation than the first candidate either because (i) it is possible to compare the fifth candidate with a previous very good candidate (contrast effect) (11, 12) or (ii) evaluators might, implicitly or explicitly, keep mental score of their evaluation during the day and try to avoid huge deviations from what they expect to be the mean evaluation of the population (narrow bracketing) (13). One crucial factor neglected in the literature is that the type of decision most commonly studied has largely been social decisions, that is, deciding about the future of other people—e.g., grading Master of Business Administration (MBA) applicants (13) or giving prisoners parole (14). This social element might also factor into the decision process. We make use of relevant research from the social decision-making literature to describe a previously unidentified mechanism in which sequence effects can affect high-stake evaluations.

If examined closely, a group of evaluators grading candidates can be understood as an interaction between two parties in which one holds all the power to decide the outcome and the other does not. This structure parallels a dictator game in which one party is given an endowment of money and can decide how much money to share with another party if any (15). In the dictator game, people are relatively generous in the first iterations of the game but become less generous as more rounds unfold (16). Repetition thus seems to erode generosity. In an evaluation setting, giving a weak candidate a pass (“sparing” them) when it is unclear whether they deserved it can be understood as an act of generosity. According to the social decision-making literature then, as the sequence unfolds, candidates will become more likely to fail if evaluators have previously acted generously. We call this the generosity-erosion effect. Here, we test this effect and compare it with the contrast and narrow-bracketing effects. Relying on data from a high-stake hiring process, we analyzed the grades of 11,281 candidates in an oral exam conducted to select public teachers in Catalonia, Spain. We show that the generosity-erosion effect has the strongest and most persuasive impact on juries' evaluations in comparison with previous effects described in the literature. Specifically, our results show that the likelihood of success decreases by 7.7% for each previous candidate that received the lowest assessment possible to continue with the hiring process. After finding prevailing evidence for the existence of the generosity-erosion effect, we test the pervasiveness of the effect by investigating

Copyright © 2021
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence RI, USA. ²Department of Applied Economics-Public Policies, University of Barcelona, Barcelona, Spain. ³Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.

*Corresponding author. Email: marclluisvives@gmail.com

its boundary conditions, which can shed light on the mechanisms behind it.

RESULTS

The generosity-erosion effect when hiring public teachers

We define as an act of generosity the evaluation of a candidate with the minimum grade possible to continue in the hiring process (5.00 of 10). Assuming that it is impossible to assess the merit of a candidate at a decimal level, which in the case of a 5.00 makes the difference between a candidate failing and passing, a final grade of 5.00 is partially a sign of the predisposition of the academic board to give a pass to an arguably weak candidate. The most frequent grade obtained in the oral exam was 5.00, probably because of an aversion to rejecting ambiguous candidates based on a decimal difference, a well-known effect in the literature (see Fig. 1) (17, 18). For our purposes, the moment in which each candidate took the oral exam was randomly determined by a candidates' surname lottery: Candidates who had a surname that started with the letter Y were called to be the first to do the exam. The list was then filled in alphabetical order. This random device successfully avoids selection biases as shown by the balance in observable characteristics (gender, age, years of experience, and grades in subsequent exams of the selection process) over our independent variable of interest (see Table 1).

On average, candidates obtained a grade of 5.50 ($SD = 2.12$), and 62% of them obtained a pass to progress to the next exams. To explore how an act of generosity affected the likelihood of passing the oral exam for subsequent candidates, we estimated linear probability model regressions with candidates' probability of passing (0 = fail, 1 = pass) as the dependent variable and the number of candidates given a 5.00 earlier on the same day by the same academic board as the treatment variable. The analysis is restricted to the fifth and later oral exams for a given day to ensure that every candidate had at least a probability above 0.10 of being part of the experimental condition, that is, that jury members previously endorsed at least one act of generosity. This restriction is applied because we considered that analyzing data in which candidates had a lower probability than 0.10 to be in the experimental condition would provide very noisy estimates of the generosity-erosion effect due to the reduced observations available of the experimental condition. Nevertheless, the result still holds whether the whole sample is analyzed, with the

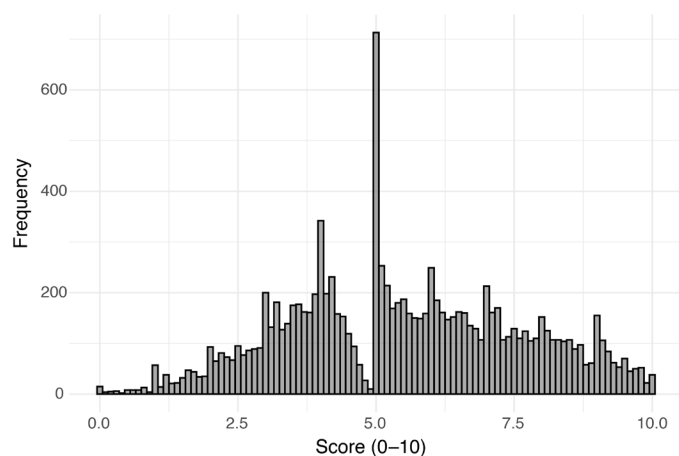


Fig. 1. Histogram of the score for hiring teachers.

effect size of the generosity-effect increasing monotonically as the probability to be in the experimental condition increases (see fig. S1).

Figure 2A shows descriptive evidence that the average grade and the candidates' probability of passing the exam decreased when candidates defended the oral exam after at least one previous candidate was graded with 5.00. Figure 2B shows the results of formally estimating the effect of generosity-erosion. We present three different specifications. In the unadjusted model, we include the number of previous candidates who obtained an exact 5.00, tribunal, day of the exam, and the order of examination as explanatory variables. In the adjusted one, we also include candidates' observable characteristics. Last, in the combined model, we also include the measure for contrast effect (i.e., a categorical variable that takes the value 1 if the prior candidate has passed the exam and 0 otherwise) and of narrow bracketing (i.e., the proportion of previous candidates who have passed the oral test). We can see that the results are very robust across specifications. In the most demanding one, the combined model, we find that the probability of passing decreases by 4.80 percentage points ($SE = 0.023$), meaning a relative probability reduction of 7.7% to pass the exam, for each previous candidate that during the day was graded with an exact 5.00. This effect is larger than the contrast and narrow-bracketing effects and the only one that remains significant in the combined model (see Table 2). Furthermore, results remain the same if the analysis is performed on grades instead of probability of passing (see table S1), showing the strength of the generosity-erosion effect, since it negatively affects all grades, not only the probability of passing (see Fig. 1A).

The analysis pipeline described was not preregistered. Researchers face several forking paths in their analytical decisions that can lead to bias (19). One method to circumvent this is to report other forking paths and still show the same result. Simply put, if the main effect survives a variety of specifications, then one can be more confident of its robustness. We conducted several analyses that operationalized the contrast and narrow-bracketing effects differently. In all cases, results remained unchanged, with the generosity-erosion effect as the only variable that significantly biased candidates' grades (see the Supplementary Materials).

Last, to control for possible effects of randomness in our data that could underly our effect, such as regression to the mean, we ran all analyses again by reshuffling the order of candidates. No effects were observed in this placebo test (see Fig. 2B), reinforcing the robustness of our results.

One crucial feature of the hiring process analyzed was that passing the exam was more likely than failing the exam (62% of candidates passed it). Because of this, evaluators might be especially akin to feelings of guilt when failing a candidate, which might lead to an avoidance to do so. Only after pardoning a failure, or several of them, evaluators will feel comfortable enough to fail ambiguous weak candidates. If this mechanism underlies the generosity-erosion effect, hiring processes that are unlikely to cause feelings of guilt will not be affected by the generosity-erosion effect. To elucidate the pervasiveness of the generosity-erosion effect and directly test this hypothesis, we explored a very selective hiring process in which failing was the norm rather than the exception.

The boundary conditions of the generosity-erosion effect: Strictness precludes lenience

Following the previous operationalization, we once again define an act of generosity as the lowest possible grade that candidates could

Table 1. Candidate's characteristics by the number of previous candidates with a score of 5.00. Categorical variables are expressed in % and continuous variables are expressed in mean (SD).

Previous 5.00 grades	None N = 2375	One N = 424	Two or more N = 58	P value
Type of position:				0.14
School	2005 (84.4%)	353 (83.3%)	47 (81.0%)	
High School	192 (8.08%)	22 (5.19%)	6 (10.3%)	
Vocational School	178 (7.49%)	49 (11.6%)	5 (8.62%)	
Gender:				0.08
Female	1945 (81.9%)	352 (83.0%)	41 (70.7%)	
Male	430 (18.1%)	72 (17.0%)	17 (29.3%)	
Hour	14.0 (2.12)	14.0 (2.19)	14.1 (2.16)	0.91
Score test 2a	5.44 (1.83)	5.54 (1.85)	5.60 (2.25)	0.67
Score test 2b	5.88 (1.90)	5.84 (1.87)	5.71 (1.83)	0.86
Past experience (years)	4.03 (4.45)	3.69 (4.35)	4.16 (4.11)	0.34

receive to continue in the hiring process. In this hiring process, in which candidates to judge were evaluated, 25.00 (of 50.00) was the cutting point (see Materials and Methods for more information). In contrast with the previous hiring process, candidates were more likely to fail than to pass the exam, as 65% of them failed. Candidates who failed were left ungraded, which precludes us from analyzing grades instead of the likelihood to pass. The same analysis pipeline conducted was applied to these data, and we preregistered before analysis. The code for all analyses, synthetic data (20) for experiment 1, as well as the entire data and preregistration for experiment 2, can be found at <https://osf.io/47ngz/>.

In the current setting, no information is available regarding the day in which each candidate was examined. Instead, we have access to the session each candidate belonged to and the number of the sequence within the session, which typically lasted for 3 days. As a consequence, the analysis and main independent variables were split by session instead of within a day. Generosity erosion was calculated by counting all the previous candidates who obtained the lowest possible grade to continue in the hiring process (i.e., 25.00 of 50). In parallel, the narrow-bracketing effect was calculated by averaging all previous candidates who passed the exam. Since the contrast effect only considered whether the previous candidate passed or not, the specification of this variable was the same as in the previous hiring process.

Results reveal that the generosity-erosion effect did not significantly change the likelihood of passing the exam, not even under the simplest specification (see Table 3). In contrast with the previous result, we find that for each SD in narrow bracketing, the probability of passing decreases by 4.88 percentage points (SE = 0.024), meaning a relative probability reduction of 14.04% to pass the exam (see Table 3). As in the previous hiring process, the contrast effect did not play a significant role in this scenario (see Table 3). Overall, when the vast majority of candidates are failing, the probability to fail does not decrease due to previous acts of generosity. However, candidates do fail more when the average of previously approved candidates deviates from what is expected, as denoted by the narrow-bracketing effect.

DISCUSSION

Collectively evaluating candidates is, ultimately, a social decision that resembles the structure of a dictator game: One party (the examining board) holds all the power to decide about the future outcome of another party (the candidates). On the basis of this insight, we linked the outcomes of an evaluation process, grades of an oral exam, with a well-known phenomenon extracted from the social decision-making literature: In dictator games, generosity is eroded by repetition (15). Results reveal that markers of past generosity within a day by the examining board (i.e., grading a candidate with the lowest grade possible to continue in the recruitment process) had a negative effect on the probability that successive candidates will pass the oral exam. Specifically, candidates were 7.7% less likely to pass for each candidate who previously obtained the lowest possible grade to continue in the process (5.00 of 10). The generosity-erosion effect was the unique predictor that prevailed after controlling for cognitive mechanisms that have been previously described as the causes of sequential effects when jurors evaluate the performance of candidates [contrast effects (11) and narrow bracketing (13)]. Furthermore, as indexed by the forking path and simulation analysis, the generosity-erosion effect is robust, since it survives several specifications and is not the by-product of randomness in the data (e.g., regression to the mean) but a structure that the examining board imposed on the data. Caution needs to be added, however, since the generosity-erosion effect in hiring decisions has not been previously demonstrated. Future research is needed to estimate its generalizability and pervasiveness in recruitment processes.

Simply put, when evaluators assessed one candidate as just good enough to be maintained in the hiring process, the cost was charged to the outcome of future candidates. Moving beyond the description of the effect, we explored its boundary conditions in a very selective hiring process. Replicating previous work, narrow bracketing significantly biased candidates' evaluation under these circumstances: A candidate was less likely to pass as the percentage of the previous candidates who passed increased. We replicated this effect in a very selective hiring process, similar to the context in which it was first described (MBA applications) (13). Our results indicate that the

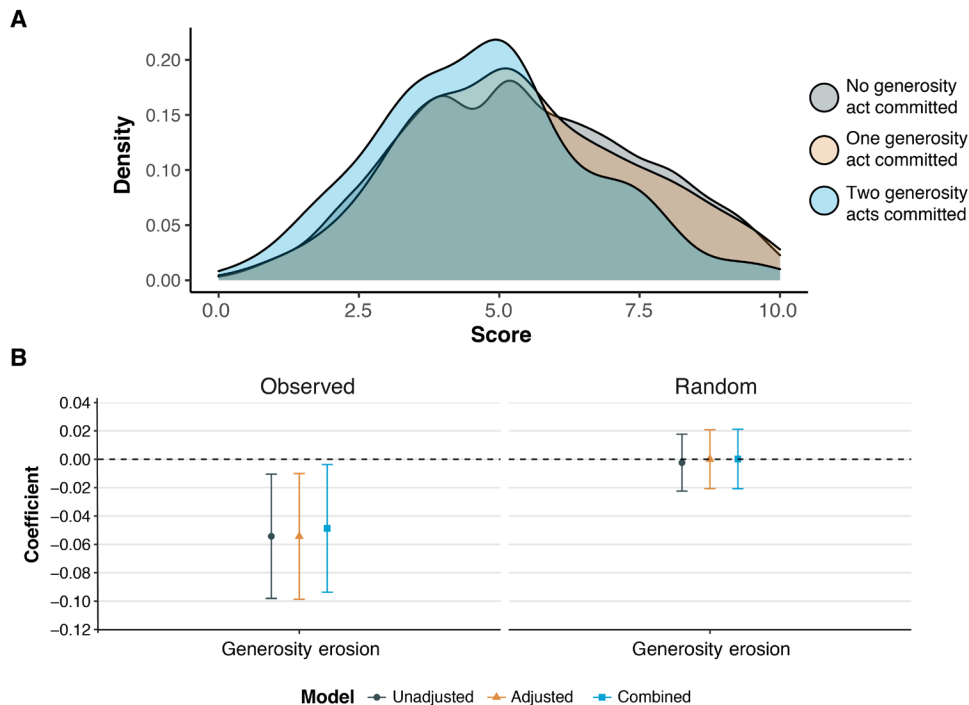


Fig. 2. Results when hiring public teachers (experiment 1). Density distribution (A) and model estimates (B) of the generosity erosion effect for hiring teachers.

Table 2. Estimates for generosity-erosion effect for hiring teachers. Estimates present the absolute probability effect. SEs were estimated with the Huber Sandwich estimator and clustered at the tribunal and day level.

Effect	Model	Estimate	95% Confidence Interval		P value
Generosity erosion	Unadjusted	-0.055	-0.098	-0.012	0.01
Contrast effect	Unadjusted	-0.038	-0.076	0.000	0.04
Narrow bracketing	Unadjusted	-0.038	-0.079	0.003	0.07
Generosity erosion	Adjusted	-0.054	-0.098	-0.011	0.01
Contrast effect	Adjusted	-0.037	-0.075	0.001	0.06
Narrow bracketing	Adjusted	-0.040	-0.081	0.002	0.06
Generosity erosion	Combined	-0.048	-0.092	-0.004	0.03
Contrast effect	Combined	-0.026	-0.069	0.017	0.23
Narrow bracketing	Combined	-0.019	-0.066	0.029	0.44

narrow-bracketing effect can also appear in recruitment processes with several evaluators instead of only one, as previously described.

Aggregated judgments are often successful at reducing individual bias and boosting accuracy (21). In the current contexts, however, the collective structure of the decision did not protect candidates from an unfairly biased assessment, regardless of the strategy used to reach a final grade. In experiment 1, we find evidence for the generosity erosion-effect despite the evaluators using a scoring-average strategy, commonly linked to improvements in judgment (22). In the same vein, we find evidence for the narrow-bracketing effect even with 10 evaluators and a grading strategy that included two steps. First, evaluators decided by majority vote if the candidate deserved to be evaluated. Second, and only if the result was affirmative, a grade was given by averaging it across the 10 evaluators. Only

when the grade was equal or more than 25 (of 50), a candidate was considered to pass the exam. The relative complexity of this grading system did not prevent the existence of biases in grading since it suffered from narrow bracketing. Future research could explore the degree to which sequential effects in collective judgments depend on the method used to aggregate judgments (23).

An open question for future investigation concerns the mechanisms causing the generosity-erosion effect. By testing the boundary conditions, we have attempted an answer while at the same time establishing the pervasiveness of the effect across hiring processes. With the current results, we propose two nonmutually exclusive psychological factors: guilt aversion and a biased evidence accumulation.

People are guilt averse: We often treat others fairly simply to avoid feelings of guilt, which are likely to arise if the expectations of

Table 3. Estimates for generosity-erosion effect for hiring judges. Estimates present the absolute probability effect. SEs were estimated with the Huber Sandwich estimator and clustered at the tribunal and day level

Effect	Model	Estimate	95% Confidence Interval		P value
Generosity erosion	Unadjusted	-0.031	-0.074	0.011	0.15
Contrast effect	Unadjusted	0.006	-0.032	0.044	0.74
Narrow bracketing	Unadjusted	-0.045	-0.087	-0.003	0.04
Generosity erosion	Adjusted	-0.028	-0.073	0.018	0.23
Contrast effect	Adjusted	0.002	-0.039	0.042	0.94
Narrow bracketing	Adjusted	-0.038	-0.082	0.007	0.09
Generosity erosion	Combined	-0.026	-0.070	0.018	0.25
Contrast effect	Combined	0.015	-0.026	0.056	0.47
Narrow bracketing	Combined	-0.042	-0.089	0.004	0.07

others are not met (24). In the current evaluative context, fairness can be defined as grading each candidate according to their merits. Guilt, thus, can emerge as the result of overgrading or undergrading a candidate. There is an intuitive asymmetry, however, between the two: Overgrading does not have a single, identifiable victim, while undergrading does have an identifiable victim. Because of it, undergrading is likely to elicit stronger feelings of guilt than overgrading (the identifiable victim phenomenon) (25). Arguably, the highest transgression is to deny a candidate the opportunity to continue in the process when they deserve it. Evaluators, therefore, are likely to be overly cautious in failing candidates to avoid type II errors, which is similar to common wisdom in society (e.g., the principle that legal systems should protect innocent people at the cost of sometimes freeing guilty people). Guilt aversion will consequently cause generosity in grading by allowing ambiguous candidates to be kept in the hiring process. However, grading several candidates with the lowest grade required to stay in the process (5.00) provides evidence, at least in the eyes of the evaluators, that they have already proven their fairness. This might reduce their feelings of guilt for failing future candidates, thus becoming harsher in their grading. An independent line of research supports this notion: People are more likely to commit a moral transgression just after proving their moral virtuosity (i.e., moral licensing (26). Simply put therefore, repetitive acts of generosity might increase harshness in grading by reducing feelings of guilt associated with them. In favor of this account, we failed to find evidence for the generosity-erosion effect in a hiring process in which evaluators were unlikely to feel guilty when failing a candidate since failing was the norm rather than the exception. High strictness in a hiring process might preclude lenience, an interesting idea that can open future avenues of research.

Alternatively, a successful model of decision making posits that there are three parameters when deciding between two options (i.e., fail/pass): A bias that makes people lean more toward one option or another, a drift rate that serves as evidence accumulation toward one of the two options, and a threshold that, once crossed, signifies that a final decision has been reached (27). If we apply this model to the scenario in which the generosity-erosion effect was observed, one explanation is that the academic board had distinct parameters for failing and for passing candidates. More candidates passed the exam than failed, suggesting that the baseline mindset for the academic board might have been to pass a candidate if not convinced

of the opposite. In contrast, this mindset is likely to be the opposite when failing is the norm rather than the exception, such as in the hiring process in which no generosity effect was observed. Grading a candidate with the lowest possible grade to continue in the process might cause a change in parameters for subsequent candidates, such as making decisions more biased toward failing, or reduce the threshold needed to fail a candidate. Sequential changes in parameters during decision making have been observed in the past, although in more basic low-level decisions such as perceptual decisions (28). The two mechanisms proposed are not mutually exclusive. One of the reasons why decisions to fail could have distinct parameters from decisions to pass might be guilt aversion.

A clear understanding of the mechanisms behind the generosity-erosion effect is crucial for policy makers since they will imply different interventions to reduce these effects in hiring processes. For example, if guilt aversion is the main mechanism underlying the generosity-erosion effect, a possible intervention could be to randomly select a subsample of the evaluators' grades to calculate the final grade. Using this protocol, none of the evaluators will know which grades lead to failure. This will probably reduce the guilt associated with failing a candidate since it has been shown that increasing uncertainty between people's actions and their outcomes causes people to behave more immorally (29), arguably because the added uncertainty alleviates feelings of guilt by allowing people to deny the role played in the negative outcome.

Similarities can be drawn between the hiring processes investigated here and past research. As previously mentioned, MBA admissions are negatively affected by the average grade of previous candidates on the same day (13). As shown here, in contexts in which receiving a positive evaluation is the exception rather than the norm, it is unlikely to find evidence for the generosity-erosion effect. On the other hand, sequential judgments of parole are a perfect candidate for the generosity-erosion effect (14). Although previous work has established sequential effects, these have been attributed to contrast with the previous candidate. Because the nature of the evaluation does not allow us to identify when evaluators were generous, it is impossible to analyze whether the generosity-erosion effect is taking place.

Companies, governments, and universities are constantly making high-stake decisions based on judgments regarding the performance and quality of candidates. Research on how people are

affected by unnecessary features when making these crucial decisions has mainly accounted for them based on cognitive mechanisms. Often, it has ignored one fundamental aspect of these decisions: They are about people. Bearing in mind the social element of these decisions, we found a novel effect that explains evaluations more powerfully than classical cognitive mechanisms previously established. Evidence taken from basic game theory paradigms is thus able to successfully provide insights on how decisions, one after another, are made.

MATERIALS AND METHODS

Experiment 1: We examined the generosity-erosion effect by analyzing data from a selection process to hire teachers in Catalonia, Spain. The Catalan Department of Education provided the data, which are available under request to the Departament d'Educació-Generalitat de Catalunya. The first part of the selection process comprised an oral exam in which each candidate defended their syllabus to an academic board of five members. After their presentation, each juror graded the candidate individually from 0 to 10. The final grade was then determined by averaging the five different grades from the academic board. If the highest and lowest grades given by the committee differed from each other by more than three points, those extreme grades were discarded and the average was calculated with the remaining three grades. This process lasted around 45 min. We use the gradings from this oral exam to investigate the presence of the generosity-erosion effect and compare it with contrast and narrow-bracketing effects (see table S1 for results regarding contrast and narrow-bracketing effects).

From each of the 20,254 candidates who enrolled to participate in this recruitment process, we obtained information on gender, years of work experience, grades in each part of the hiring process (the oral exam and two subsequent written exams), and their area of specialty (among 47). Furthermore, we can identify each of the 182 examining boards evaluating the candidates. After discarding candidates who did not present themselves to the oral exam, we analyze the data from the resulting 11,281 candidates.

Experiment 2: We replicated the analysis using data from candidates who are in the hiring process to become judges in Spain. The data are available on the website of the Spanish General Council of the Judiciary. This hiring process consisted of three exams. The first one was multiple choice with 100 questions. Only a portion of candidates is defined as the best and make it through to the next stage (around 1500 candidates/year). The second exam, which is the one we analyzed, consists of an oral presentation (60 min). Specifically, candidates must orally explain five topics each, randomly assigned the day of the exam for each candidate, and assessed by an examining board of nine members. After the presentation, the board casts a vote to accept or reject the candidate. If the candidate gets the majority of affirmative votes, then they evaluate whether they obtained a grade higher or equal to 25.00 points (of 50 points). Only under this circumstance, the candidate remains in the selection process.

We analyzed the calls in 2018 and 2019 of the Spanish Judiciary. A total of 8546 (4193 in 2018 and 4353 in 2019) candidates took part in the selection process and only 3171 passed the first exam (1651 in 2018 and 1520 in 2019). Candidates were then allocated along 35 sessions and 6 boards per call for the oral exams. Each session lasted 3 days, with 16 candidates per session on average. For each candidate, we obtained information regarding the grading in the

multiple choice exam and the oral exam, their examining board, the examining session, and the order in which they presented within the session.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/17/eabe2045/DC1>

REFERENCES AND NOTES

1. A. Rubinstein, Y. Salant, A model of choice from lists. *Theor. Econ.* **1**, 3–17 (2006).
2. W. B. de Bruin, Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychol.* **118**, 245–260 (2005).
3. V. A. Ginsburgh, J. C. Van Ours, Expert opinion and compensation: Evidence from a musical competition. *Am. Econ. Rev.* **93**, 289–296 (2003).
4. J. Levav, M. Heitmann, A. Herrmann, S. S. Iyengar, Order in product customization decisions: Evidence from field experiments. *J. Polit. Econ.* **118**, 274–299 (2010).
5. N. Augenblick, S. Nicholson, Ballot position, choice fatigue, and voter behaviour. *Rev. Econ. Stud.* **83**, 460–480 (2015).
6. D. Feenberg, I. Ganguli, P. Gaulé, J. Gruber, It's good to be first: Order bias in reading and citing NBER working papers. *Rev. Econ. Stat.* **99**, 32–39 (2017).
7. E. K. Kelley, P. C. Tetlock, How wise are crowds? Insights from retail orders and stock returns. *J. Finance* **68**, 1229–1265 (2013).
8. B. Mellers, L. Ungar, J. Baron, J. Ramos, B. Gurcay, K. Fincher, S. E. Scott, D. Moore, P. Atanasov, S. A. Swift, T. Murray, E. Stone, P. E. Tetlock, Psychological strategies for winning a geopolitical forecasting tournament. *Psychol. Sci.* **25**, 1106–1115 (2014).
9. J. Becker, D. Brackbill, D. Centola, Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5070–E5076 (2017).
10. F. Galton, Vox populi. *Nature* **75**, 450–451 (1907).
11. S. Bhargava, R. Fisman, Contrast effects in sequential decisions: Evidence from speed dating. *Rev. Econ. Stat.* **96**, 444–457 (2014).
12. S. M. Hartzmark, K. Shue, A tough act to follow: Contrast effects in financial markets. *J. Finance* **73**, 1567–1613 (2018).
13. U. Simonsohn, F. Gino, Daily horizons: Evidence of narrow bracketing in judgment from 10 years of M.B.A. admissions interviews. *Psychol. Sci.* **24**, 219–224 (2013).
14. S. Danziger, J. Levav, L. Avnaim-Pesso, Extraneous factors in judicial decisions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6889–6892 (2011).
15. D. Kahneman, J. L. Knetsch, R. H. Thaler, Fairness and the assumptions of economics. *J. Bus.* **59**, S285–S300 (1986).
16. P. D. Bó, G. R. Fréchet, On the determinants of cooperation in infinitely repeated games: A survey. *J. Econ. Lit.* **56**, 60–114 (2018).
17. R. Diamond, P. Persson, The long-term consequences of teacher discretion in grading of high-stakes tests. *National Bureau of Economic Research*, No. 22207 (2016); 10.3386/w22207.
18. T. S. Dee, W. Dobbie, B. A. Jacob, J. Rockoff, The causes and consequences of test score manipulation: Evidence from the New York regents examinations. *Am. Econ. J. Appl. Econ.* **11**, 382–423 (2019).
19. A. Gelman, E. Loken, The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Dep. Stat. Columbia Univ.* **348** (2013).
20. D. S. Quintana, A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *eLife* **9**, e53275 (2020).
21. J. Surowiecki, *The Wisdom of Crowds* (Anchor, 2005).
22. D. Ariely, W. T. Au, R. H. Bender, D. V. Budescu, C. B. Dietz, H. Gu, T. S. Wallsten, G. Zauberman, The effects of averaging subjective probability estimates between and within judges. *J. Exp. Psychol. Appl.* **6**, 130–147 (2000).
23. D. Prelec, H. S. Seung, J. McCoy, A solution to the single-question crowd wisdom problem. *Nature* **541**, 532–535 (2017).
24. P. Battigalli, M. Dufwenberg, Guilt in games. *Am. Econ. Rev.* **97**, 170–176 (2007).
25. K. Jenni, G. Loewenstein, Explaining the identifiable victim effect. *J. Risk Uncertain.* **14**, 235–257 (1997).
26. I. Blanken, N. van de Ven, M. Zeelenberg, A meta-analytic review of moral licensing. *Pers. Soc. Psychol. Bull.* **41**, 540–558 (2015).
27. R. Ratcliff, P. L. Smith, A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* **111**, 333–367 (2004).
28. R. Ratcliff, P. L. Smith, S. D. Brown, G. McKoon, Diffusion decision model: Current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
29. J. Dana, R. A. Weber, J. X. Kuang, in *Economic Theory* (2007).

Acknowledgments: We thank the Departament d'Educació of Generalitat de Catalunya (Gabinet Tècnic) for providing the data used in this article. We also thank to M. Gonzalez-Avià, L. Farré, L. González, J. Jofre-Monseny, and F. Ortega for providing valuable input; A. Drew for

valuable comments on previous versions of the manuscript; and J. van Baar for proposing the final title of the article. **Funding:** This work was supported by grants from the Spanish Ministerio de Ciencia e Innovación PID2019-110397RA-I00, by Secretaria General de Recerca-Generalitat de Catalunya SGR2017-644 and AGAUR (FI-DGR). **Author contributions:** All authors contributed equally to this work. **Competing interests:** M.S.-B. declares conflict of interest grants from the European Commission H2020 programme and the EIT Health programme, outside the submitted work. All other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The code for all analyses, synthetic data (20) for experiment 1, as well as the entire data and

preregistration for experiment 2, can be found at <https://osf.io/47ngz/>. Additional data related to this paper may be requested from the authors.

Submitted 19 August 2020

Accepted 4 March 2021

Published 21 April 2021

10.1126/sciadv.abe2045

Citation: M.-L. Vives, T. Fernandez-Navia, J. J. Teixidó, M. Serra-Burriel, Lenience breeds strictness: The generosity-erosion effect in hiring decisions. *Sci. Adv.* **7**, eabe2045 (2021).