Outcome bias in self-evaluations: Quasi-experimental field evidence from Swiss driving license exams

Meier, Pascal Flurin ; Flepp, Raphael ; Meier, Philippe ; Franck, Egon

DOI: https://doi.org/10.1016/j.jebo.2022.07.013

Originally published at:
Meier, Pascal Flurin; Flepp, Raphael; Meier, Philippe; Franck, Egon (2022). Outcome bias in self-evaluations: Quasi-experimental field evidence from Swiss driving license exams. Journal of Economic Behavior Organization, 201:292-309.
DOI: https://doi.org/10.1016/j.jebo.2022.07.013

Journal of Economic Behavior and Organization 201 (2022) 292–309

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Outcome bias in self-evaluations: Quasi-experimental field evidence from Swiss driving license exams

Pascal Flurin Meier*, Raphael Flepp, Philippe Meier, Egon Franck

University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 3 March 2022
Revised 8 July 2022
Accepted 12 July 2022

JEL classification:
D81
D83
D91

Keywords:
Outcome bias
Self-evaluation
Behavioral economics
Judgment
Regression discontinuity design

ABSTRACT

Exploiting a quasi-experimental field setting, we examine whether people are outcome biased when self-evaluating their past decisions. Using data from Swiss driving license exams, we find that candidates who narrowly passed the theoretical driving exam are significantly less likely to pass the subsequent practical driving exam – which is taken several months after the theoretical exam – than those who narrowly failed. Those candidates who passed the theoretical exam on their first attempt receive more objections regarding their momentary, on-the-spot decisions in the practical exam, consistent with the idea that the underlying behavioral difference is worse preparation.

1. Introduction

Do people weigh the outcome of a decision more heavily than its informational content when making self-evaluations? For example, consider the decision to drive a car under the influence of alcohol. The driver might evaluate this decision more positively after an accident-free ride than after a crash under otherwise similar circumstances. Consequently, the driver's future decisions regarding intoxicated driving might be distorted by the outcome of this ride. Such behavior is referred to as outcome bias (Baron and Hershey, 1988).

Prior empirical evidence on outcome bias has primarily stemmed from third-party evaluation settings, in which an evaluator assesses the decision of an agent (e.g., Brownback and Kuhn, 2019; Gauriot and Page, 2019; Rubin and Sheremeta, 2016). In contrast, evidence on outcome bias in self-evaluations is scarce and almost exclusively based on laboratory experiments (Jones et al., 1997; Ratner and Herbst, 2005; Tinsley et al., 2012). One notable exception is the field study by Lefgren et al. (2015), who found that professional basketball coaches exhibit outcome bias when revising their starting lineup decisions. However, it remains unclear whether their results are driven by outcome bias in the self-evaluations of the coaches or by outcome bias among third-party evaluators (e.g., general managers, fans, or the media) and with unbiased coaches responding accordingly to retain their legitimacy. Overall, even though self-evaluations account for the majority of individuals' real-life decision-making, a thorough understanding of whether outcome bias in self-evaluations distorts subsequent decisions in the field is still lacking.

* Corresponding author.
E-mail address: pascal-flurin.meier@business.uzh.ch (P.F. Meier).

https://doi.org/10.1016/j.jebo.2022.07.013

We fill this research gap by investigating the behavior of young adults in the context of Swiss driving license exams. This setting offers several unique advantages for studying outcome bias in self-evaluations. First, the Swiss driving licensing system comprises one computer-based theoretical exam and one practical exam on the streets separated by several months, which allows us to test whether outcome bias in self-evaluations after the first exam impacts the results of the second exam. Second, the candidates represent the entire population of young adults in Switzerland because almost everyone participates in the driving license process, enabling us to investigate outcome bias in a broad population.

Finally, the threshold for passing the theoretical car driving exam (hereafter, the "theoretical exam") allows us to compare two groups of candidates who performed very similarly but with one group barely passing the exam and the other barely failing. Conditional on the number of errors in the theoretical exam around the passing threshold, the outcome of either passing or failing is uninformative about the candidate's exam performance. Thus, this setting generates a quasi-experiment in which the only difference between the candidates around the passing threshold is the exam outcome. If the candidates exhibit outcome bias in their self-evaluations, we would expect to observe a difference in the probability of passing the practical car driving exam (hereafter, the "practical exam") between the two groups.

Using data on more than 40,000 candidates, we employ a regression discontinuity design (RDD) to test whether self-evaluations and thus the probability of passing the practical exam differ between candidates barely failing and those barely passing the theoretical exam. Our results suggest that the candidates who barely failed the theoretical exam on the first attempt are approximately 6 percentage points more likely to pass the practical exam on their first attempt than candidates who barely passed the theoretical exam on their first attempt. This finding is robust to various validation and falsification checks. Thus, the candidates who barely failed the theoretical exam seem to evaluate their own performance more negatively than candidates who barely passed, leading the former to increase their effort in preparing for the practical exam relative to the latter.

Despite its advantages, our setting suffers from a potential concern about the interpretation of the results. Because the candidates who barely failed the first theoretical exam need to retake that exam before being allowed to take the practical exam, they may simply accumulate more theoretical knowledge, which helps them on the practical exam. To distinguish between these explanations, we make use of our rich dataset and draw from the literature on procedural knowledge (Knowlton et al., 2017; Sanchez and Reber, 2013) and skill acquisition (Lewin, 1982). While retaking the theoretical exam might increase declarative knowledge, the procedural knowledge of how to drive a car can be gained only through experience and practice. Thus, if outcome bias leads to a different response in terms of the decision regarding how much to prepare for the practical exam, we expect the candidates who barely failed the theoretical exam on their first attempt to perform better in the evaluation categories on the practical exam that demand more procedural knowledge.

Indeed, we find that the candidates who barely failed the theoretical exam receive significantly fewer objections in the evaluation categories "visual" and "tactics", which are more closely related to momentary, on-the-spot decisions that require procedural knowledge. However, we fail to find significant differences in the evaluation categories that are more closely related to declarative knowledge such as "handling" or "maneuvers". These results suggest that the candidates who experienced a narrow failure prepare themselves better for the practical exam than candidates who experienced a narrow success. This pattern supports the outcome bias explanation and contradicts the alternative explanation regarding the accumulation of theoretical knowledge.

We add to the literature in several ways. First, our study provides novel field evidence on outcome bias in self-evaluations. In contrast to Lefgren et al. (2015), we find that the source of outcome bias in our setting likely stems from the candidates because third parties such as driving instructors or parents have only limited knowledge about the candidates' performance on their theoretical exam and their preparation level. In addition, we show that outcome bias has consequences, while Lefgren et al. (2015) focus on changes in the decision-maker's strategy. Second, we are the first to document outcome bias in the general population using data that almost perfectly capture the young Swiss population. This approach thus extends the previous field evidence that uses sports data and involves a subpopulation of individuals acting in a high-stakes environment (e.g., Gauriot and Page, 2019; Kausel et al., 2019; Lefgren et al., 2015). Finally, due to its quasi-experimental nature, our setting allows for a clean interpretation of the effect of the outcome information on subsequent performance as causal.

Our results have important practical implications. First, outcome bias in self-evaluation is present in the (young) population as a whole and is thus not limited to certain subpopulations. Second, there are many similar institutional settings with multiple separated exams, such as most educational environments or professional licensing settings. Thus, outcome bias in self-evaluations might substantially impact the career paths of individuals. Finally, to mitigate the adverse consequences of outcome bias, candidates could use debiasing strategies (Soll et al., 2014). Plausible solutions might be to sensitize candidates to the fact that the result "barely passing the exam" could have easily been the result "barely failing the exam".

The remainder of this paper is structured as follows. In Section 2, we discuss the related literature. In Section 3, we describe the research design. The results are presented in Section 4, and in Section 5, we address one major alternative explanation. Section 6 concludes the paper with a discussion.

2. Related literature

First labeled by Baron and Hershey (1988), outcome bias refers to the phenomenon in which people place excessive weight on the importance of outcome information in decision-making. Outcomes should be considered in an evaluation only if they provide additional information on the quality of the decision. For instance, if a decision-maker has more information than does a judge, then outcomes may be a "valid, albeit imperfect indicator of decision quality" (Hershey and Baron, 1992, p. 90). However, whenever a judge has the same information as a decision-maker ex ante, outcomes are uninformative and should be ignored (Hershey and Baron, 1992). Thus, people are outcome biased when a behavior is considered more justifiable in light of a favorable outcome (Lefgren et al., 2015).

In laboratory experiments, Baron and Hershey (1988) examine subjects who evaluate the appropriateness of third-party decisions, although they also discuss outcome bias in the context of self-evaluation.[1] For instance, their experiments include assessing the appropriateness of a cure (e.g., surgical operation) that had fixed, ex ante probabilities (e.g., 8%) of causing severe consequences (e.g., death). The authors consistently show that the subjects rated the decisions significantly better and the decision-maker as more competent when the outcome was favorable than they did when it was unfavorable.

Most of the subsequent work in laboratory experiments focuses on individuals' assessing third-party behaviors—such as military combat (Lipshitz, 1989), legal (Alicke et al., 1994) or financial decisions (König-Kersting et al., 2021), ethical judgments (Gino et al., 2010), salesperson performance (Marshall and Mowen, 1993), or audit quality (Peecher and Piercey, 2008) or stems from other principal agent settings and games (e.g., Brownback and Kuhn, 2019; Cushman et al., 2009; Gurdal et al., 2013). Outcome bias is even present in third-party evaluations in which the evaluator has complete information on the underlying decision-making process, for instance, when he or she knows the investment strategy of the agent (König-Kersting et al., 2021) or observes effort (Brownback and Kuhn, 2019).

Field evidence on outcome bias in third-party evaluations mainly stems from sports settings. A notable exception is Emerson et al. (2010), who find a positive outcome bias in peer reviews of evidence-based medicine. Both Gauriot and Page (2019) and Kausel et al. (2019) examine outcome bias utilizing data from professional soccer. Gauriot and Page (2019) exploit the quasi-arbitrary outcome of shots that hit the goal post. It is as good as random whether a long-distance shot that hits the post ultimately lands inside or outside the goal. Thus, the outcome of a long-distance shot to the post does not provide information about the quality of the shot. However, the authors find evidence that lucky post-in shots result in more playing time for the shooter in the subsequent match and in higher ratings by both journalists and fans than do unlucky post-out shots. In a similar vein, Kausel et al. (2019) examine narrow winnings from penalty shootouts in professional football. The authors find that players on the winning team receive higher journalistic ratings, suggesting that these experts are overly influenced by the outcome of the game. This is true even for players who did not participate in the penalty shootout and therefore could not actively contribute to the outcome of the game after regular play and overtime.

In contrast to evidence from third-party evaluations, evidence is scarce on outcome bias in self-evaluations of the appropriateness of decisions. However, these kinds of decisions are highly relevant since they are common in everyday life and have direct personal consequences. In addition, Lefgren et al. (2015) argue that it is more significant to identify outcome bias in self-evaluations given that people find it easier to be dismissive of others' decisions than of one's own. Unlike external judges who act rationally when relying on outcome information in the absence of other information (Baron and Hershey, 1988), individuals who self-evaluate possess process knowledge. They have perfect knowledge about the information available to the decision-maker at the time of the decision and about the specific process underlying the decision (Jones et al., 1997). Given these conditions, outcome information is not indicative of decision quality, and any effect that outcome information has on evaluations of one's own decisions is evidence of bias (Jones et al., 1997).

In the context of self-evaluations, Jones et al. (1997) show that outcome information not only influences participants' evaluations of their decision-making process but also affects their memories in a way that is consistent with subsequent outcomes. The authors suggest that individuals who experience a good outcome recall their decision-making process as being more thoughtful. Ratner and Herbst (2005) find that a negative emotional response resulting from the unfavorable outcome of a good decision may lead individuals to switch to inferior alternatives. Bachmann (2018) shows that advisers can eliminate outcome bias in the context of investment decisions, particularly after bad outcomes. However, the advisers are unable to prevent affective reactions after bad outcomes and instead might even reinforce them. Examining decisions in the context of a hazardous situation, Tinsley et al. (2012) find that outcome bias also affects future risk-taking. In addition to conducting laboratory experiments, the authors administer field surveys to residents who assess their actual evacuation decisions following hazard situations. Individuals with resilient near-miss experiences of disasters underestimate the danger of future hazardous situations and increase the risk in their decision-making (e.g., not engaging in activities to mitigate the potential hazard).

The only field study investigating self-evaluation is Lefgren et al. (2015), who propose that professional coaches exhibit outcome bias when revising their strategies. Utilizing data from top-tier basketball teams, the authors employ an RDD to study how coaches react to close outcomes that are uninformative about team effectiveness or future success and therefore should not impact the coaches' strategies. However, Lefgren et al. (2015) find that coaches are more likely to revise their starting lineup after a narrow defeat than they are after a narrow win, suggesting that outcome bias is present in the high-stakes decision-making of professionals. The authors acknowledge that they cannot rule out the possibility that the source of the outcome bias is another stakeholder group, i.e., a third party, since coaches must appease a broad set of constituencies (e.g., the general manager, fans, and media). Thus, the actions of the coaches might be perfectly rational.

[1] Baron and Hershey (1988, p. 578) note that people who judge their behaviors as a function of their outcomes "may hold themselves responsible for both good and bad luck, becoming smug in their success or self-reproachful in their failure".

Ultimately, outcome bias can have severe consequences for both the evaluated agents and the evaluators. Tinsley et al. (2012) suggest that the interpretation of an outcome may lead to different subsequent behaviors in risky decision-making contexts, resulting in major monetary and emotional costs when an unfavorable outcome occurs. In the sports context, Lefgren et al. (2015) suggest that excessive strategy switching among coaches matters: It may lead to worse outcomes in future games and impose direct costs, such as placing an emotional strain on the actors involved. Flepp and Franck (2021) examine coach dismissal decisions and find that decisions based on factors beyond the coaches' control have major consequences. The authors show that only dismissals following actual poor performance on the pitch improve subsequent team performance, while dismissals after seemingly poor performance (e.g., due to bad luck) do not. Similarly, in the context of CEO turnover decisions, Flepp (2021) shows that corporate boards consider uninformative outcomes outside of the control of the CEO in their decision-making.

In summary, there is a broad stream of literature addressing outcome bias in different contexts. However, evidence on self-evaluation is sparse and predominantly stems from laboratory experiments. In the field, outcome bias in self-evaluations has been examined only in the high-stakes setting of professional team sports (Lefgren et al., 2015). However, given the natural environment of team sports and its broad set of constituencies, a clear attribution of outcome bias to a specific stakeholder group and thus a generalization to self-evaluations is challenging. In the subsequent section, we carefully document how we address outcome bias in the context of self-evaluations within the general population of young adults when applying for a Swiss driving license.

3. Research design

3.1. Institutional setting and data

We utilize the novel setting of young adults applying for driving licenses in Switzerland. According to the Federal Statistical Office of Switzerland (2022), in 2015, 82% of the Swiss population (8.3 million at that time) was allowed to drive a car. Since a great majority of the Swiss population participates in the Swiss driving license process at some point, our data closely mirrors the general population of individuals in their 20s. This is a distinct advantage of our data.

Adults are allowed to drive motor vehicles after they have successfully passed a theoretical and a practical exam. The theoretical exam is a computer-based multiple-choice exam with 50 questions.[2] A candidate successfully passes the exam if he or she makes fewer than 16 errors. If a candidate fails the exam, he or she can retake it after scheduling a new appointment and paying a registration fee. The agency issues a learner's license valid for 24 months if the exam is passed. This license allows the holder to drive a car with an eligible adult as a codriver and to attend driving lessons with a driving instructor. If the student driver feels ready, he or she can register for the practical exam after attending an additional sensitization course consisting of 8 teaching hours.[3] The practical exam involves up to 60 min of driving with an independent official driving expert acting as the codriver. The expert judges the student driver according to a defined set of criteria in different evaluation categories. If the student driver fails the exam, he or she can retake it in the near future.

Switzerland consists of 26 states, called cantons. The driving license process is consistent across the cantons, but each canton is responsible for issuing driving licenses to the citizens of that canton. The data was provided by three cantons—A, B, and C—which we keep anonymous. Our sample period starts in 2014 and runs until 2018. We focus on car exams only and on applicants who have no prior driving experience (e.g., a scooter license). We have detailed data at the individual level, which enables us to merge the candidates' theoretical exam results with their practical exam result. We start with an initial sample of 57,238 candidates for whom we have data on the theoretical and the first practical exam. We retain all the observations for whom we have full data on their first theoretical exam, data on subsequent attempts if they failed the exam initially, and the result of their first practical exam. After cleaning the data, the resulting sample includes 44,486 observations with full data on the theoretical and practical exam outcomes.[4]

For the theoretical exam, we observe the number of errors on the first attempt (*Errors*). *FirstAttemptFailed* is a dummy variable equaling 1 if the candidate failed their first attempt at the theoretical exam (i.e., *Errors* is greater than or equal to 16 errors) and 0 otherwise. We have additional information on gender (*Gender* is a dummy equaling 1 if the candidate is male and 0 otherwise), nationality (*Swiss* is a dummy equaling 1 if the candidate is Swiss and 0 otherwise), and age (*Age* measures the candidate's age at the time of the theoretical exam).

For our outcome variable, performance on the practical exam, we have rich data. Most importantly, we observe whether a candidate passed the exam (*PractExam* is a dummy equaling 1 if the exam is passed and 0 otherwise). Furthermore, we have detailed information about why the examinant failed the exam. The examiner assesses the candidate based on several factors. If the examiner objects to an action by the candidate, he or she categorizes the violation according to a predefined catalog of more than 40 possible objections (see Table A1 in the Appendix for the complete list). If the candidate commits

[2] The theoretical exam can be scheduled one month prior to the applicant's 18th birthday at the earliest. Applicants who have a driving license in a different category (e.g., motorbike) do not have to take the theoretical exam. In our sample, we include only applicants taking the theoretical exam for the first time.

[3] Typically, the driving instructor enrolls the student; however, the candidates may also self-register.

[4] To ensure data validity, we deleted the observations for whom we have strong evidence of flaws (e.g., exams labeled as passed with 1 or more objections or duplicate observations).

Table 1
Descriptive statistics.

	Definition	N	Mean	Q1	Median	Q3	SD
Dependent variables							
PractExam	1 if the candidate passed practical exam on first attempt	44,486	0.642	0.000	1.000	1.000	0.479
#Objections	Number of objections	44,486	3.317	0.000	0.000	7.000	4.868
Independent variables							
FirstAttemptFailed	1 if the candidate failed theoretical exam on first attempt	44,486	0.140	0.000	0.000	0.000	0.347
Errors	Number of errors on the first theoretical exam	44,486	8.449	3.000	6.000	11.000	8.064
Age	Age of the candidates when taking the theoretical exam	44,486	21.318	18.142	18.764	21.493	5.788
Gender	1 if the candidate is male	44,486	0.487	0.000	0.000	1.000	0.500
Swiss	1 if the candidate is Swiss	44,486	0.787	1.000	1.000	1.000	0.410

Notes: This table reports the descriptive statistics for the variables employed. N denotes the number of observations. Mean denotes the mean of the corresponding distribution. The table also displays the medians, the 25th percentile (Q1), the 75th percentile (Q3), and the standard deviations (SD) of the distribution.

one or more violations, the candidate fails the exam. The candidates who pass the practical exam have zero objections, while those who fail the exam receive at least one objection. We calculate the number of individual objections (*#Objections*) as an alternative outcome variable.

In Table 1, we present the descriptive statistics for our employed variables. A total of 64% of the candidates pass the practical exam on their first attempt, and the average number of objections is approximately 3.3. Regarding the theoretical exam, 86% of candidates make 15 or fewer errors and thus pass the exam on their first attempt. We plot a histogram of *Errors* on the first theoretical exam within the window of 0 to 30 errors in Fig. A1 in the Appendix. Most of the candidates make relatively few errors, with a smooth gradual decrease in the number of errors. The average number of errors on the theoretical exam during the first attempt is approximately 8.5. The average time between passing the theoretical exam and the first practical exam attempt is 247 days, i.e., approximately eight months. In terms of the demographic characteristics of our sample, the average candidate is approximately 21 years old, 48.7% of the candidates are male, and 78.7% of the candidates are Swiss.

In the context of our institutional setting, failing the theoretical or practical exam has both monetary and nonmonetary consequences. Candidates who fail the theoretical exam pay a relatively small fee (approximately USD 38) to register for a retake, which takes the average candidate approximately 37 days to pass. In addition, candidates who retake the exam may bear nonnegligible reputational costs, particularly since the theoretical exam is considered no major hurdle, given the passing rate of 86% on the first attempt.

Regarding failing the practical exam, the costs are substantially higher. On the one hand, the candidates must wait to be able to drive a car (approximately 50 days according to our data), which, for people at this age, is an important life event and reputational signal. On the other hand, failing also generates direct monetary costs. The monetary cost of an additional attempt is at least USD 142 when no additional driving lessons are needed but is more likely to reach approximately USD 870, which includes five additional driving lessons and the provision of the driving instructor's car for the practical exam.[5] The majority of young Swiss adults who take the practical exam are typically in their final year of vocational training, and earn a median gross income of USD 1,360 per month (Federal Statistical Office, 2021). Thus, failing the practical exam imposes nonnegligible costs for these young adults.

Overall, the incentives to pass the practical exam on the first attempt seem to be higher than those for the theoretical exam. However, compared to other field studies investigating outcome bias (e.g., Gauriot and Page, 2019; Kausel et al., 2019; Lefgren et al., 2015), in our setting, the stakes are generally much lower.

3.2. Empirical strategy

We employ an RDD to estimate the causal effect of different outcomes on future performance. Our quasi-experiment exploits an arbitrary cutoff in the running variable, which we use to assign individuals to the treatment or control conditions. This approach relies on the assumption that the characteristics of the candidates vary smoothly through the threshold. In this case, any discontinuity in future performance may be causally attributed to the treatment, i.e., experiencing a narrow negative outcome.

We test for outcome bias by examining whether the candidates who barely failed or barely passed the theoretical exam exhibit different probabilities of passing the practical exam, conditional on their first theoretical exam attempt. While a candidate experiences a positive outcome when narrowly passing the theoretical exam (e.g., passing with 15 errors), a candidate experiences a negative outcome in the case of a narrow failure (e.g., failing with 16 errors). Thus, both the candidates

[5] The underlying assumptions are as follows (CHF 1 = USD 1.09 as of October 26, 2021): *Administrative costs:* USD 142: Fee for the practical exam (Federal Chancellery, 2022). *Driving lessons:* A lesson price depends on the canton and ranges between USD 87 to 120 (Federal Chancellery, 2022). We take the average (USD 104). USD 520: Costs of five additional driving lessons (excluding lessons with family and friends). We assume one lesson every 10 days given the average time interval of 50 days between practical exam attempts. USD 208: Costs of two additional driving lessons on the day of the practical exam (incl. using the driving instructor's car for the practical exam).

receive an informative signal about their performance, i.e., the number of errors, suggesting barely sufficient preparation for the exam, and they receive a conditionally uninformative signal of performance, i.e., the positive or negative outcome of the exam. Crucially, the two candidate types exhibit almost identical performance, suggesting that their ex ante strategic choices were very similar (e.g., they invested a similar amount of time in preparation). If candidates near the threshold have different passing probabilities for the subsequent practical exam, it indicates outcome bias in self-evaluations.

To empirically examine the candidates' future performance after experiencing different outcomes, we rely mainly on three sets of results. We first graphically investigate the presence of a discontinuity around the threshold. To this end, we plot the local sample means of our dependent variable in the nonoverlapping bins of the number of errors on the theoretical exam. Second, we quantitatively estimate our RDD. We rely on a local linear regression, i.e., a nonparametric approach, which is frequently used in empirical RD analyses due to its good balance between flexibility and simplicity (Cattaneo et al., 2019).[6] We test the robustness of our baseline choice and estimate the regression using the whole sample, i.e., using a parametric approach. Third, we perform a similar estimation with more detailed performance information to more precisely identify the mechanism at work.

Similar to Klein Teeselink et al. (2021), we use the local linear method proposed by Calonico et al. (2014a) to strike an appropriate balance between bias and precision associated with the use of a local or global RDD approach. Calonico et al. (2014a) implement a nonparametric local polynomial estimation method with an optimal bandwidth selection and robust confidence intervals.[7] For an in-depth discussion of this methodology, refer to Calonico et al. (2014b) and Calonico et al. (2019). For the empirical implementation, we utilize the Stata command *rdrobust* developed by the authors (Calonico et al., 2014a; Calonico et al., 2017). Our baseline RD models correspond to the following local regression model in which the observations are weighted by a kernel function[8] and the bandwidth is both data-driven and mean squared error (MSE) optimal:

$$PractExam_i = \alpha + \gamma_0 f(Errors_i - \bar{e}) + \beta FirstAttemptFailed_i + \gamma_1 f(Errors_i - \bar{e}) FirstAttemptFailed_i + \delta X_i + \mu_i \tag{1}$$

We are interested in estimates of the coefficient β, which equals the treatment effect, i.e., the effect of failing the theoretical exam, on future performance. $PractExam_i$ is a dummy variable indicating whether the candidate successfully passed the subsequent practical exam. f denotes a suitable polynomial function of $Errors_i$, which indicates the number of errors on the theoretical exam, centered at the threshold for passing ($\bar{e}$).[9] We allow for potentially different slopes on either side of the threshold; thus, the coefficients on the polynomial terms are indexed by 0 and 1. X_i denotes a vector of covariates. We estimate our baseline specification both with and without covariates. We control for predetermined characteristics such as gender (*Gender*), age (*Age*), and Swiss citizenship (*Swiss*) and include canton fixed effects. Conceptually, covariates are not necessary in an RDD and serve mainly to increase precision (Calonico et al., 2019).

We also estimate Eq. (1) using a logistic regression over the entire candidate sample, corresponding to the parametric approach. Conceptually, the parametric approach equals the nonparametric approach using a uniform kernel with the maximum bandwidth (Cattaneo et al., 2019).

To evaluate our baseline results, we conduct a battery of validity tests. First, we check the sensitivity of the RD estimates to different bandwidths, kernel functions, and polynomial orders. Second, we perform balance checks to examine whether the treatment and control groups are similar in terms of observable characteristics near the cutoff. Third, we examine the treatment effects at placebo cutoffs to test whether the regression function is continuous at thresholds other than the actual treatment cutoff. Finally, we estimate several alternative specifications.

4. Results

4.1. Main results

In this section, we present the results of our empirical estimation. We start by illustrating the RDD graphically. In Fig. 1, we plot the mean likelihood of passing the practical exam by the number of errors on the theoretical exam and fit a second-order polynomial using a broad window of 0 to 30 errors. In Fig. 2, we focus more narrowly on the local area around the threshold of 16 errors and plot only observations with between 10 and 20 errors. Furthermore, we fit a local

[6] Following the literature, our preferred model uses a first-order polynomial (Cattaneo et al., 2019; Gelman & Imbens, 2019). Higher-order polynomials tend to overfit the data and lead to unreliable results near boundary points, which, together with the increasing variability of the treatment effect estimator, cause the local linear estimator to be the preferred point estimator in many applications (Cattaneo et al., 2019).

[7] The selection of the optimal bandwidth is data-driven. That is, it is determined on the basis of a nonparametric approximation that is the result of a trade-off between lower variance (associated with a larger bandwidth) and higher bias (associated with the poorer parametric polynomial approximation when using a larger bandwidth) (Calonico et al., 2014a). The authors correct for the misspecification of the confidence intervals as a consequence of the larger bandwidths. They provide a new theory-based and more robust confidence interval estimator for average treatment effects at the cutoff using a bias-corrected RD estimator together with a novel standard error estimator.

[8] We employ a triangular kernel in the baseline model, which places a positive but declining (linearly and symmetrically) weight on observations within the bandwidth and is zero otherwise. When used in conjunction with a bandwidth that optimizes on the basis of the mean squared error (MSE), it results in a point estimator with ideal properties (Cattaneo et al., 2019).

[9] For the sake of symmetry in the bandwidth around the threshold, we utilize a threshold of 15.5 errors. We also check the results using a threshold of 16 errors, which leads to similar results.

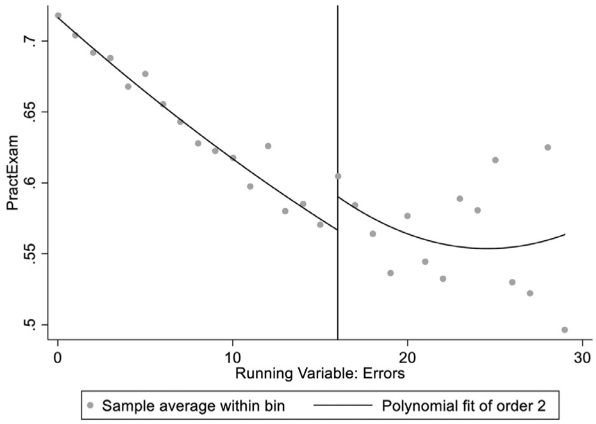

Fig. 1. RD Plot 0–30 Errors.
Notes: The figure shows the RD plot for candidates who had between 0 and 30 errors. The line is obtained by fitting a second-degree polynomial that is allowed to differ on either side of the cutoff.

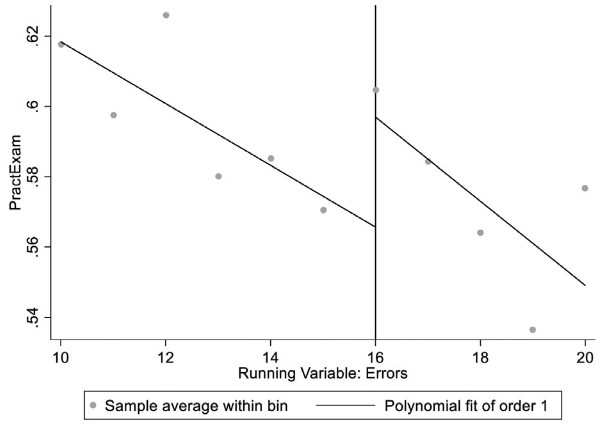

Fig. 2. RD Plot 10–20 Errors.
Notes: The figure shows the RD plot for candidates who had between 10 and 20 errors. The line is obtained by fitting a first-degree polynomial that is allowed to differ on either side of the cutoff.

first-order polynomial, which closely resembles our baseline RD estimation. Both figures show a clearly visible discontinuity around the passing threshold, hinting towards a positive treatment effect from experiencing a narrow failure: Candidates who experienced a narrow failure on the theoretical exam seem to be more likely to pass the subsequent practical exam on their first attempt than candidates who narrowly passed the theoretical exam.

We next estimate the magnitude of the effect of experiencing a narrow failure relative to that of experiencing a narrow pass using both a nonparametric and a parametric approach. The baseline estimates from the nonparametric approach are reported in Table 2, Panel A. In addition to the treatment coefficients and the standard errors in parentheses, we report the number of observations to the left (#L) and right (#R) of the cutoff and the estimation bandwidth for each model in the nonparametric approach. We obtain both conventional RD estimates with a conventional variance estimator and a bias-corrected RD estimate with a robust variance estimator, as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017).

Model I in Panel A of Table 2 suggests that candidates who narrowly failed their theoretical exam but pass it on a subsequent attempt are more likely to pass the practical driving exam on their first attempt. We interpret this evidence as being consistent with outcome bias. Depending on the estimator, the candidates who barely failed the theoretical exam on their first attempt are 5.1 (conventional RD estimator with conventional standard errors) to 6.0 (bias-corrected estimator with robust standard errors) percentage points more likely to successfully pass the practical driving exam on their first attempt than candidates who barely passed the theoretical exam. Table 2, Panel A, Model II includes the candidate characteristics as the covariates, which should not greatly affect our estimates if the RDD assumptions are valid (Calonico et al., 2019). Indeed, the coefficients remain virtually unchanged.

In Panel B of Table 2, we present the estimates from the parametric approach including all of our sample observations and using a logistic regression model.[10] We report the baseline results both with and without additional covariates. Again, the results show that the estimated treatment coefficient is positive and statistically significant, suggesting that the candidates who narrowly failed the theoretical car driving exam are more likely to pass the subsequent practical driving exam.[11]

Overall, the results suggest that relative to narrowly passing the theoretical exam, narrowly failing the theoretical exam leads to a considerable increase in the probability of passing the subsequent practical driving exam. These results are consistent with the idea of outcome bias. One plausible explanation is that the two groups of candidates differ in their level of preparation. For instance, candidates who narrowly passed the theoretical exam may not prepare as much as candidates who barely failed since they feel more justified in their ex ante decision. We address this potential underlying mechanism in more detail in Section 5.

4.2. Validity and robustness tests

4.2.1. Sensitivity to choice of bandwidth, kernel, and polynomial

Local polynomial estimation requires the researcher to make different implementation decisions regarding the bandwidth, the kernel function, and the order of the polynomial. In what follows, we test the model's sensitivity to our baseline choices.

First, we use different bandwidths, as suggested by Calonico et al. (2017). We start with a window of 13 to 18 errors and gradually increase it to 8 to 23 errors. Second, we check the robustness of the baseline results to the use of different kernel

[10] The results are robust to the use of a linear probability model.

[11] Our results are also insensitive to the additional inclusion of examiner fixed effects.

Table 2
Baseline results.

Dependent Variable: *PractExam*		
Panel A: Nonparametric approach	**(I)**	**(II)**
Beta (Conventional)	0.051**	0.055**
	(0.024)	(0.025)
Beta (Robust)	0.060**	0.064**
	(0.029)	(0.030)
Bandwidth	4.697	4.524
#L	6,492	6,492
#R	2,911	2,911
Covariates	0	5
Panel B: Parametric approach	**(I)**	**(II)**
Beta	0.118***	0.100**
	(0.044)	(0.045)
Observations	44,486	44,486
R-Squared	0.01	0.03
Covariates	0	5

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses.
Panel A: Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). The model is estimated using a triangular kernel and includes a first-degree polynomial, which is allowed to differ on either side of the cutoff. Model II includes the following covariates: *Age, Swiss, Gender*, and canton fixed effects.
Panel B: The sample includes all observations in our sample (parametric approach) weighted by a uniform kernel. The model is estimated using a logistic regression and includes a first-degree polynomial, that is allowed to differ on either side of the cutoff. Model II includes the following covariates: *Age, Swiss, Gender*, and canton fixed effects.

functions. We employ a uniform kernel, which places equal weight on all observations within the bandwidth, and a Epanechnikov kernel, which places a quadratic decaying weight on observations within the bandwidth (Cattaneo et al., 2019). Finally, we test for sensitivity to the polynomial order by utilizing higher-order polynomials than our default choice of a first-order polynomial. Again, we report both the conventional RD estimates with a conventional variance estimator and the bias-corrected RD estimates with a robust variance estimator.

The results are displayed in Table 3. Our estimates are mostly robust to changes in the window, at least in terms of the size of the treatment effect, though statistical significance is not present for the narrow windows of 13 to 18 and 12 to 19 errors. Moreover, our results are not sensitive to the choice of kernel function, while they are somewhat sensitive to polynomials of order three, at least in terms of statistical significance. Overall, these sensitivity tests show that the RD estimates are stable and remain mostly statistically significant. These results support the validity of our findings.

4.2.2. Balance checks

One of the most important RD falsification tests is to examine whether the treatment and control groups are similar in terms of observable characteristics near the cutoff (Cattaneo et al., 2019). If candidates lack the ability to precisely manipulate the number of errors that they make, those who narrowly passed the theoretical driving license exam are likely to be similar—except for their exam outcome—to those who narrowly failed the exam in terms of their observable (e.g., age) and unobservable (e.g., preparation for the theoretical exam) characteristics.

This assumption is feasible for several reasons. First, the candidates have no prior testing experience at this level, given that the theoretical exam involves a completely new task and a completely new context. Thus, they are unlikely to predict exactly the preparatory amount needed to (narrowly) pass the theoretical exam. Second, similar to Proud (2015), even if candidates could predict the amount of effort needed, their realized errors are still expected to vary unpredictably around the threshold.

Table 3
Sensitivity to choice of bandwidth, kernel, and polynomial.

Dependent Variable: *PractExam*						
Panel A: Bandwidth	(I)	(II)	(III)	(IV)	(V)	(VI)
Beta (Conventional)	0.055	0.047	0.056**	0.045**	0.039*	0.035*
	(0.037)	(0.029)	(0.025)	(0.023)	(0.021)	(0.019)
Beta (Robust)	0.064	0.064	0.041	0.066*	0.063**	0.058**
	(0.054)	(0.054)	(0.041)	(0.036)	(0.032)	(0.029)
Kernel	Trian.	Trian.	Trian.	Trian.	Trian.	Trian.
Polynomial Order	1	1	1	1	1	1
Window	13–18	12–19	11–20	10–21	9–22	8–23
#L	3,359	4,867	6,492	8,315	10,437	12,834
#R	1,974	2,481	2,911	3,282	3,605	3,875
Covariates	5	5	5	5	5	5
Panel B: Kernel function	(I)	(II)	(III)	(IV)	(V)	(VI)
Beta (Conventional)	0.051**	0.055**	0.061**	0.062**	0.056**	0.056**
	(0.024)	(0.025)	(0.024)	(0.024)	(0.025)	(0.025)
Beta (Robust)	0.060**	0.064**	0.069**	0.073***	0.064**	0.066**
	(0.029)	(0.030)	(0.027)	(0.028)	(0.030)	(0.030)
Kernel	Trian.	Trian.	Uni.	Uni.	Epa.	Epa.
Polynomial Order	1	1	1	1	1	1
Bandwidth	4.697	4.524	4.297	3.727	4.419	4.225
#L	6,492	6,492	4,867	4,867	4,867	4,867
#R	2,911	2,911	2,481	2,481	2,481	2,481
Covariates	0	5	0	5	0	5
Panel C: Polynomial degree	(I)	(II)	(III)	(IV)	(V)	(VI)
Beta (Conventional)	0.051**	0.056**	0.060	0.055**	0.058**	0.063
	(0.024)	(0.028)	(0.040)	(0.025)	(0.028)	(0.038)
Beta (Robust)	0.060**	0.063*	0.059	0.064**	0.066**	0.060
	(0.029)	(0.032)	(0.047)	(0.030)	(0.032)	(0.043)
Kernel	Trian.	Trian	Trian	Trian	Trian	Trian
Polynomial Order	1	2	3	1	2	3
Bandwidth	4.697	7.834	7.802	4.524	7.768	8.509
#L	6,492	12,834	12,834	6,492	12,834	15,521
#R	2,911	3,875	3,875	2,911	3,875	4,135
Covariates	0	0	0	5	5	5

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). In Panel B and C, the model includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). The model is estimated using the indicated kernel and polynomial function, which is allowed to differ on both sides of the cutoff. The following covariates are included if indicated: *Age*, *Swiss*, *Gender*, and canton fixed effects.

To mitigate any remaining concerns, we use the same local polynomial regression model, i.e., Eq. (1) using different predetermined covariates as the dependent variables. As suggested by Cattaneo et al. (2019), we analyze each covariate as if it were an outcome. Table 4 shows no evidence of preexisting differences in the covariates between the treatment and control groups. Thus, the validity of the RDD is further supported.

4.2.3. Treatment effect for placebo cutoffs

The key identifying assumption relies on the continuity of the regression function for the treatment and control observations at the cutoff in the absence of treatment. Unfortunately, this condition is fundamentally untestable. However, a useful falsification test for the RDD, in addition to the covariate balance checks, is to examine the treatment effects at artificial or placebo cutoffs. Thus, we test whether the regression function is continuous at thresholds other than the actual treatment cutoff. Evidence of continuity apart from the actual threshold is not a necessary or a sufficient condition for continuity at the actual cutoff. However, discontinuities away from the cutoff would raise questions of validity of the RDD (Cattaneo et al., 2019).

Thus, we examine the neighboring artificial cutoffs of 14, 15, 17, and 18 errors to check for discontinuities at locations other than the actual treatment cutoff of 16 errors. In addition, we use the artificial cutoffs of plus-minus 5 errors, i.e., 11 and 21, to check for discontinuities even further away from the actual cutoff. The results are presented in Table 5. We find no evidence of any discontinuities using the artificial cutoffs, which leaves us confident that our identification strategy is valid.

Table 4
Balance checks.

Dependent Variable: Predetermined Covariates			
	(I) Age	(II) Gender	(III) Swiss
Beta (Conventional)	0.137	−0.037	0.023
	(0.252)	(0.026)	(0.017)
Beta (Robust)	0.196	−0.047	0.026
	(0.302)	(0.031)	(0.021)
Bandwidth	6.542	4.361	7.986
#L	10,437	4,867	12,834
#R	3,605	2,481	3,875

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). The model is estimated using a triangular kernel and includes a first-order polynomial, that is allowed to differ on either side of the cutoff.

Table 5
Alternative cutoffs.

Dependent Variable: *PractExam*							
	(I) 11	(II) 14	(III) 15	(IV) 16 (Actual)	(V) 17	(VI) 18	(VII) 21
Beta (Conventional)	−0.012	−0.007	0.004	0.051**	−0.001	−0.027	−0.020
	(0.021)	(0.022)	(0.021)	(0.024)	(0.023)	(0.022)	(0.031)
Beta (Robust)	−0.020	−0.010	−0.000	0.060**	0.005	−0.033	−0.033
	(0.025)	(0.027)	(0.026)	(0.029)	(0.028)	(0.027)	(0.035)
Bandwidth	3.378	4.514	5.208	4.697	5.999	7.814	6.472
#L	6,342	8,357	7,344	6,492	7,256	9,743	3,882
#R	4,412	4,054	3,452	2,911	2,841	2,918	1,618
Covariates	0	0	0	0	0	0	0

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported, as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). The model is estimated using a triangular kernel and includes a first-order polynomial, that is allowed to differ on either side of the cutoff.

4.2.4. Alternative specifications

To further substantiate the robustness of our results, we check several alternative specifications. First, we utilize the number of objections (*#Objections*) as an alternative outcome variable. If a candidate passes the practical exam, this variable equals zero. Thus, under this specification, we expect a negative effect. The results displayed in Table 6 are robust under both estimators and are in line with the baseline results, indicating that candidates who experience a narrow failure on the theoretical exam receive fewer objections on the practical exam. Thus, we confirm that the candidates who failed narrowly at first not only have a higher probability of successfully passing the practical exam but also receive fewer objections.

Second, even though we find no evidence of covariate imbalance around the threshold (see Section 4.2.2), we perform additional robustness checks in which we match candidates in terms of covariates. We match candidates who narrowly failed the theoretical exam with candidates who narrowly passed the theoretical exam based on their predetermined characteristics, i.e., based on exact matches in terms of *Gender, Swiss, Age,* canton of residence, and distance to the cutoff. Thus, candidates with 15 errors are matched to candidates with 16 errors, while candidates with 14 errors are matched to candidates with 17 errors and so on. After matching, we have an equal number of (similar) candidates who passed the theoretical exam and who failed the exam on their first attempt.

In our first analysis, we depart from the RDD and compare the average outcomes on the practical exam between matched candidates who narrowly passed (failed) the theoretical exam with 15 (16) errors. The results of a simple *t* test are displayed in Table A2 in the Appendix. The difference in the proportion of candidates who narrowly failed the theoretical exam with 16 errors (61.2% of candidates) and candidates who narrowly passed the theoretical exam with 15 errors (56.5% of candidates)

Table 6
Alternative dependent variable.

Dependent Variable: *#Objections*		
	(I)	(II)
Beta (Conventional)	−0.476*	−0.551**
	(0.255)	(0.253)
Beta (Robust)	−0.566*	−0.651**
	(0.307)	(0.304)
Bandwidth	4.954	4.738
#L	6,492	6,492
#R	2,911	2,911
Covariates	0	5

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). The model is estimated using a triangular kernel and includes a first-order polynomial, that is allowed to differ on either side of the cutoff. Model II includes the following covariates: *Age, Swiss, Gender*, and canton fixed effects.

is statistically significant. This result provides supportive evidence for our previous conclusions without relying on any RDD implementation choices.

In our second analysis, we reestimate our baseline RD specification using only the matched treatment and control sample. This process ensures that the number of candidates to the left and to the right of the cutoff are exactly the same. Again, the RD estimates remain significant and comparable in size (the results are not reported for brevity).

As a final robustness test, we conduct a different RD approach as proposed in Cattaneo et al. (2015) and Cattaneo et al. (2017) and implemented by Cattaneo et al. (2016), which is based on the ideal of local randomization. The results are displayed in Table A3 in the Appendix and remain similar

5. Alternative explanation

Our baseline results suggest that the candidates who narrowly pass the theoretical exam are more likely to pass the subsequent practical exam. This result is consistent with the explanation of outcome bias. However, the institutional setting also provides a potential alternative explanation. Since the candidates who fail their first theoretical exam must retake the theoretical exam before being allowed to take the practical exam, those who barely failed may simply have accumulated more theoretical knowledge, which helps them on the practical exam, than those who barely passed the theoretical exam.

To test this alternative explanation empirically, we make use of our rich data on the individual objections during the practical exam, categorized according to the official classification used by the Road Traffic Department in each canton. The objection categories are handling, vision, environment, tactics, dynamics, maneuvers, and control. If knowledge accumulation is the main mechanism, we would expect to observe a different pattern among the candidates who barely failed or passed the theoretical exam.[12]

Each objection category includes multiple individual objections.[13] Cantons A and B utilize the same classification list, and Canton C uses a similar but slightly different list.[14] A full list of the individual objections as well as their categorizations and corresponding short descriptions is presented in Table A1 in the Appendix. We consider objection categories (e.g., maneuvers) that subsume multiple individual objections (e.g., parking, reversing, or turning).

To theoretically connect the objections to the accumulation of knowledge as an alternative explanation, we draw on the literature on procedural knowledge and learning (Knowlton et al., 2017) and on the acquisition of skills (Lewin, 1982).

[12] We also attempt to control for the effect of the retake exam itself by including the number of errors on the retake exam as an additional covariate in the RD analysis. The RD estimates remain statistically significant and even increase slightly in magnitude.

[13] For instance, if a candidate fails to properly park in reverse, incorrectly uses the emergency brake, or overlooks a right of way, the examiner marks objections in parking, emergency brake, and right of way, respectively. The candidate fails the practical exam as he or she has three objections. Each individual objection is classified under one broader category; while parking and emergency brake are considered incorrectly performed maneuvers, a right of way objection falls under the traffic tactics category.

[14] The categories for these objections are the same in all three cantons. As an exception, cantons A and B include the additional category “miscellaneous” (a pool of objections that are not classified). While the majority of the objections overlap, some objections are unique or are classified differently. For instance, cantons A and B have only one objection for parking, while canton C differentiates among parking sideways, forward and in reverse.

Table 7
Objections.

Dependent Variable: Indicator Variable for the Objection Category							
	(I) Maneuvers	(II) Visual	(III) Dynamics	(IV) Environ.	(V) Handling	(VI) Tactics	(VII) Control
Beta (Conventional)	−0.022	−0.074***	−0.026	−0.002	−0.002	−0.062**	−0.033
	(0.021)	(0.025)	(0.020)	(0.011)	(0.016)	(0.025)	(0.023)
Beta (Robust)	−0.029	−0.085***	−0.034	−0.006	−0.006	−0.072**	−0.040
	(0.025)	(0.030)	(0.024)	(0.013)	(0.019)	(0.030)	(0.028)
Bandwidth	4.663	4.192	5.868	5.160	5.182	4.410	4.563
#L	6,492	4,867	8,315	6,492	6,492	4,867	6,492
#R	2,911	2,481	3,282	2,911	2,911	2,481	2,911
Covariates	5	5	5	5	5	5	5

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses. Conventional RD estimates with a conventional variance estimator and bias-corrected RD estimates with a robust variance estimator are reported as suggested by Calonico et al. (2014b) and implemented by Calonico et al. (2017). The sample includes observations within the optimal bandwidth selected by a common MSE-optimal bandwidth selector (Calonico et al., 2017). The model is estimated using a triangular kernel and includes a first-order polynomial, that is allowed to differ on either side of the cutoff. All the models include the following covariates: Age, Swiss, Gender, and canton fixed effects.

Learning how to drive a car is an example of the acquisition of a complex perceptual-motor skill (Fitts, 1964, 1965; Fitts and Posner, 1967; Fleishman, 1965; Lewin, 1982).[15] In the associative stage of skill acquisition (i.e., refinement and coordination through experience and practice), procedural knowledge, representing procedures learned through experience (e.g., accelerate when merging onto a highway) or “knowing how”, is acquired (Knowlton et al., 2017). This contrasts with declarative memory, or “knowing that” (e.g., knowledge of the speed limit). The concern comes down to the extent to which retaking the theoretical exam and the associated accumulation of declarative knowledge directly translate into superior performance on the practical exam.

If our results are driven by the accumulation of theoretical knowledge, we expect to observe a stronger effect among objections that are more closely associated with declarative knowledge than with procedural knowledge. We expect declarative knowledge to be more important for objections classified as maneuver or car handling objections. These objections require knowledge of how the operating equipment works, how to position the vehicle, or how to perform a maneuver that is not subject to time pressure such as parking in reverse, which follows the steps covered in a typical learning manual. Conversely, if the results are indeed driven by outcome bias, the amount of time invested in preparation for the practical exam likely differs between the two groups of candidates. This should translate into a distinct pattern of objections in which candidates who barely failed the theoretical exam receive fewer objections in categories that are more closely associated with experience.[16]

Examining the impact of attention on sensorimotor skills in novices and experts, Beilock et al. (2002) find that the performance of novices is impaired in environments that divert attention away from primary tasks, whereas more experienced performers are not adversely affected. In contrast, novices actually benefit from conditions that direct their attention toward the task properties, while experienced agents are harmed when they must devote explicit attention to skill processes that normally run automatically. As procedural knowledge develops through practice, skills that rely on such knowledge can be executed with less deliberate attention and become more readily accessible in distracting circumstances. This phenomenon may help the candidates direct their attention to other stimuli and better handle the conditions that create dual-task environments (Beilock et al., 2002), such as those often encountered on the road.

We expect such dual-task environments to be particularly present in the visuals, dynamics, and tactics categories. These categories require substantial attentional resources since they relate to dynamics on the road, such as interactions with other road users, which rely on hands-on driving experience. In our empirical implementation, we generate an indicator variable equal to 1 if the candidate violates at least one criterion in an objection category and 0 otherwise. Thus, we construct the indicator variables *Maneuvers, Visual, Dynamics, Environment, Handling, Tactics,* and *Control.* Finally, we fit the same baseline estimation, i.e., Eq. (1), using these indicator variables as the outcome variable in seven separate RD specifications.[17]

The results in Table 7 suggest that the candidates who experienced an initial setback are significantly less likely to receive objections in the visual and tactical categories, which are more closely related to hands-on experience on the road.

[15] Candidates start by acquiring declarative knowledge (“the cognitive stage”), which is assessed through the theoretical driving exam. Then, they proceed to actually learning how to drive while building up procedural knowledge (“the associative stage”).

[16] Ideally, we would observe the preparation time and effort, such as, for example, the number of driving lessons taken by each candidate, for both groups. However, we do not have access to this information. Although the data allow us to measure the interval between the passed theoretical exam and the subsequent practical exam, we opt to focus on the pattern of objections since the interval between the two is likely to be an ambiguous proxy for preparation effort. On the one hand, one could expect candidates who failed their theoretical exam to have a longer interval because they take more driving lessons. On the other hand, one could also expect candidates who had to retake the theoretical exam to be faster if they prepare more intensively or want to catch up while their counterparts slack. The data suggests that candidates who narrowly failed the theoretical exam tend to take the practical exam approximately 10 to 15 days sooner than candidates who narrowly passed it. This difference is not statistically significant in our baseline RD estimation without covariates but is significant in our baseline RD estimation with covariates.

[17] Our results are similar when we consider the number of objections per category.

In contrast, the candidates do not differ in the probability to receive objections in the maneuver and control categories, which are more closely related to declarative knowledge.[18] Thus, the results support the idea that the candidates who experienced a narrow failure invested more in preparation than the candidates who experienced a narrow success. Overall, this pattern is consistent with the outcome bias explanation and contradicts the alternative explanation of the accumulation of theoretical knowledge as the main driver of our findings.

6. Discussion and conclusion

We present novel field evidence suggesting that individuals who self-evaluate their behavior exhibit outcome bias. We find that the candidates who narrowly passed the theoretical car driving exam on their first attempt have a significantly lower probability of passing the practical car driving exam than those who barely failed the theoretical exam. Examining more detailed performance data for the practical exam, we find that the candidates who narrowly failed the theoretical exam receive fewer objections related to their momentary, on-the-spot decisions, suggesting that they might be better able to manage dual-task environments due to greater experience. We interpret these findings as being consistent with the notion of outcome bias. Our results suggest that candidates who passed the theoretical exam on their first attempt and candidates who failed the theoretical exam on their first attempt are likely to differ in the amount of time invested in preparation for the practical exam.

In the context of our institutional setting, outcome bias leads to both monetary and nonmonetary consequences. The monetary costs of failing the practical exam are at least USD 142 but more likely sum to approximately USD 870, which is bearable but not negligible for the average household in Switzerland and is a substantial amount for young adults. Furthermore, the candidates face reputational costs from failing the practical exam and must wait to finally drive a car independently, which is a significant life event for the typical young adult. In addition, the candidates who fail the practical exam face opportunity costs since they must spend additional time improving their driving skills. Thus, while most of the existing field studies on outcome bias address the decision behavior but not its consequence, our results suggest that outcome bias in self-evaluations leads to monetary and nonmonetary costs.[19]

One plausible mechanism by which outcome bias may affect the provision of effort and ultimately outcomes on the practical exam is biased belief updating relative to the Bayesian benchmark (e.g., Bühren and Krabel, 2019; Kieren and Weber, 2020). Kieren and Weber (2020) propose and find supporting evidence in the laboratory that individuals form more optimistic (negative) beliefs after observing desirable (undesirable) but uninformative signals. Similarly, Murad and Starmer (2021) find that relative performance feedback may create biases in confidence levels, even if the feedback is completely uninformative. Given that a narrow success on the theoretical exam is a desirable but uninformative signal of exam performance, while a narrow failure is an undesirable but uninformative signal, outcome bias may lead candidates who barely passed the theoretical exam to form more (or even overly) optimistic beliefs than candidates who barely failed. As a consequence of their biased beliefs, the former group might prepare less extensively for the practical exam.

The link between past performance, beliefs, and subsequent decision-making has been addressed in different settings. For instance, evidence suggests that past success may induce overconfidence, which can distort the decision-making of nonprofessional traders (Czaja and Röder, 2020), professional investors (Puetz and Ruenzi, 2011), sell-side financial analysts (Hilary and Menzly, 2006), and even CEOs (Billett and Qian, 2008; Hilary and Hsu, 2011). In the context of professional basketball, Bühren and Krabel (2019) suggest that players who experience success, i.e., who score an equalizer before overtime, exhibit a “sloppiness effect”: Their performance during overtime is substantially worse than their usual performance, which is consistent with the explanation of player overconfidence affecting subsequent performance.

Alternatively, narrowly falling short of a goal might have a motivational effect on individuals. Thus, the candidates who narrowly failed the theoretical exam might increase their effort level. Several studies have documented that people work harder if they are slightly behind in terms of reaching a goal than when they have already reached it (e.g., Allen et al., 2017; Pope and Simonsohn, 2011). For instance, in the context of basketball, Berger and Pope (2011) find that narrowly losing before halftime leads to a discontinuously higher probability of winning the game.[20] Corroborating lab experiments

[18] One institutional detail might further support the claim that the candidates are comparable in terms of declarative knowledge: Before being allowed to take the practical exam, all the candidates must attend an additional basic theoretical sensitization course consisting of eight lessons.

[19] As an exception, Lefgren et al. (2015) indicate the consequences from the outcome-biased behaviors of coaches who excessively change their strategy following a narrow defeat using a simulation approach.

[20] However, in an extensive analysis of different sports, Klein Teeselink et al. (2021) conclude that Berger and Pope's (2011) findings do not generalize across sports contexts.

in Berger and Pope (2011) demonstrate that an increase in effort provision among the trailing individuals is the causal mechanism, suggesting that being slightly behind can increase success by increasing motivation.

Our study complements this literature on performance after a success or failure by investigating how uninformative prior outcomes are linked to subsequent performance. However, one limitation of our institutional setting is that we cannot cleanly observe the behavioral adjustments made by the candidates who passed (failed) the theoretical exam narrowly. Instead, we observe only their subsequent outcomes. By examining detailed performance data and relying on the theory of the acquisition of skills and procedural knowledge, we try to minimize the concern that the alternative explanation of the accumulation of theoretical knowledge underpins the main effect.

Relatedly, we cannot infer the extent to which either or both groups adjust their strategy. We observe only the relative difference between the two groups, which suggests that at least one group has fallen prey to outcome bias. It may be that candidates who barely failed the theoretical exam increase their preparation effort due to motivational effects, that candidates who barely passed the theoretical exam decrease their preparation effort due to overconfidence, or both.

We believe that there is room for future research on outcome bias in the context of self-evaluations. Since it is crucial to understand how to enhance decision-making through strategies other than simply raising awareness, examining the factors that accentuate or attenuate outcome bias in self-evaluations could be a fruitful avenue for further research. As a notable exception in the context of self-evaluations, Bachmann (2018) shows that advising can eliminate outcome bias by reducing uncertainty in the quality of decisions, particularly after bad outcomes. Regarding third-party evaluations, Sezer et al. (2016) find that outcome bias is reduced when participants evaluate an individual separately relative to when participants evaluate individuals jointly. Furthermore, Gillenkirch and Velthuis (2018) show that outcome bias in third-party evaluations is stronger when peer comparisons are present, i.e., when a principal can compare his or her outcome with the outcomes of his or her peers. However, in the context of self-evaluations in which subjects have perfect knowledge about the underlying decision-making process, the factors that might moderate outcome bias remain to be further explored.

In conclusion, our study involving the population of young adults in Switzerland suggests that outcome bias in self-evaluations is likely to be a rather general tendency among individuals. Since many similar institutional settings have multiple sequential exams (e.g., most educational environments and professional licensing settings), outcome bias in self-evaluations might substantially impact the career paths of individuals.

We assume that agents in other economic contexts may also exhibit outcome bias when evaluating their own decisions, such as people reconsidering their retirement saving strategies, amateur sportsmen assessing their performance, or adults underestimating the risks of driving a car while texting or being under the influence of alcohol following an accident-free ride. Thus, given the numerous tasks involving self-evaluations that people face each day, we suspect that outcome bias might be common in everyday decision-making.

Declaration of Competing Interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

The authors thank the editor and two anonymous referees for their very constructive feedback and suggestions during the review process, which substantially improved the paper. The authors also thank seminar participants at the University of Zurich for their helpful comments. In particular, we owe gratitude to Lauren Howe for providing valuable suggestions.

Appendix

Table A1
Objection overview.

Objection	Description	Canton	Category
First steps	Ignition, checking in all directions, brake test	A/B/C	Handling
Operating equipment	Frame, lights, blinkers, odometer, etc.	A/B/C	Handling
Positioning	Wheel, seat, headrest, mirrors	A/B/C	Handling
Basic operations	Acceleration, use of different brakes, use of clutch, shifting, gear selection	A/B/C	Handling
Familiarity	Driver comfort, passenger seating	A/B/C	Handling
Viewing systematics	Order of checks, double checking and looking to the side, follow-up of observations, appropriateness of checks	A/B/C	Visual
Foresight	Recognition of blind spots, driving behavior, defensive driving	A/B/C	Visual
Sensoring	Road users, weather, road conditions	A/B/C	Visual
Viewing technique	Checking around curves, at bottlenecks, around bends, and at branches	A/B/C	Visual
Orientation	Front, back and side mirrors	A/B/C	Visual
Partnering	Follows the “3As”, communication with and consideration of other drivers	A/B/C	Environ.
Conditions	Awareness of road conditions, visibility and weather conditions	A/B/C	Environ.
Speed	Traveling with the flow of traffic, differentiation of appropriate speeds, speed adjustments, excessive speed	A/B/C	Dynamics
Movement	Communication, lane	A/B/C	Dynamics
Road use	Following road partitions, staying in the lane, driving on the right	A/B/C	Dynamics
Curves	Following leftward curves, following rightward curves, cutting inside turns, following the appropriate sequence of actions, lane selection, speed	A/B/C	Dynamics
Changing lanes	Use of gaps, turning left/right, following the appropriate sequence of actions, checking surroundings, navigating obstructions, control of the vehicle	A/B/C	Dynamics
Passing	Crossing into the other lane, passing, overtaking other vehicles, distance between vehicles, navigating obstructions, time, speed, allowed	A/B/C	Dynamics
Keeping up	Column driving, side by side, following other cars, distance between cars	A/B/C	Dynamics
Partnering maneuvers	Passing, use of gaps, traveling with the flow of traffic, acceptable highway driving	C	Dynamics
Driving physics	Stopping, smoothness of driving	C	Dynamics
Roundabouts	Navigation, dynamics, use of gaps	A/B/C	Tactics
Signaling	Presence, timing	A/B/C	Tactics
Entering roads	Sequence, with/without acceleration lane, entering one-way roads, crossing lanes, crossing tram tracks	A/B/C	Tactics
Braking readiness	Presence, timing, and appropriateness of braking readiness	A/B/C	Tactics
Right of way	Safety, disregard of right of way, wavering in decisions, signaling, yielding to those from the right, uncertainty, quality of handling	A/B/C	Tactics
Traffic signals	Uncertainty, disregard for signals, following signals, respecting signals (traffic signals, signposts, traffic lights, road markings, police, traffic regulation)	A/B/C	Tactics
Pedestrians	Right of way, behavior toward pedestrians, disregard of pedestrians	A/B/C	Tactics
Public transport	Behavior around buses, trams, and railway crossings	A/B	Tactics
Highway	Entry and exit, speed, distance between cars, overtaking other cars, driving on the right	A/B/C	Tactics
Obstruction	Obstruction	A/B/C	Control
Hazards	Dealing with abstract hazards, likely hazards, and concrete hazards	A/B/C	Control
Intervention	Verbal, steering, braking, or accelerating intervention needed	A/B/C	Control
Ending exam	Exam ended early	A/B/C	Control
Stopping/starting	Gradualness of stopping/starting, starting/stopping on a slope, observation of surroundings when starting/stopping	A/B/C	Maneuvers
Parking	Parking on the left, parking on the right, pulling in to park, 90-degree back-up parking, parking in reverse and parallel parking; observation of surroundings; respect for right of way; making appropriate corrections	A/B/C	Maneuvers
Reversing	Turned the wrong direction, observation of surroundings, respect for right of way, ensured sufficient space, appropriate speed, uncertainty	A/B/C	Maneuvers
Turning	Choice of place to turn, appropriateness of choice to turn, observation of surroundings, respect for right of way	A/B/C	Maneuvers
Securing	Space, sequence, usage of wedge	A/B/C	Maneuvers
Helpers	Use of support persons	A/B	Maneuvers
Ramp	Appropriate spacing, correcting, sideways, backward	A/B/C	Maneuvers
Motorbike course	Ability to drive straight down the lane, to slalom, and to drive in a figure-8, insufficient ability to maneuver, uncertainty, driving too fast, aborted attempted maneuver; crashed	A/B/C	Maneuvers
Emergency brake	Insufficient use of emergency brake, unsafe use of emergency brake	A/B/C	Maneuvers

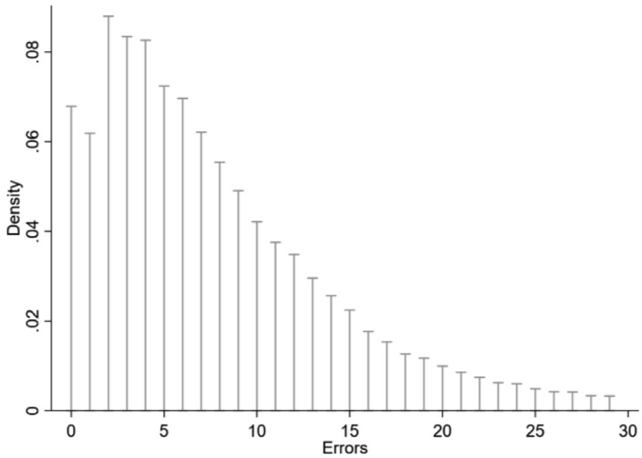

Fig. A1. Error histogram.
Notes: Errors distribution between 0 and 30.

Table A2
Matching approach.

t-test Group	N.	Mean	Std. Error
FirstAttemptFailed = 0	619	0.565	0.020
FirstAttemptFailed = 1	619	0.612	0.020
Diff.		−0.047	0.028
Ha: diff < 0		Pr (T < t) = 0.047**	
Ha: diff!= 0		Pr (|T| > |t|) = 0.094*	

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. t-test for equality in means of *PractExam* by *FirstAttemptFailed* for the candidates with 15 and 16 errors.

Table A3
RDD under local randomization.

Dependent Variable: *PractExam*						
Panel A: First-order polynomial	(I)	(II)	(III)	(IV)	(V)	(VI)
Diff. in means	0.034	0.052***	0.043***	0.061***	0.034***	0.032***
p value	0.150	0.000	0.000	0.000	0.002	0.000
Bandwidth	1	2	3	4	5	6
#L	971	2,080	3,359	4,867	6,492	8,315
#R	764	1,428	1,974	2,481	2,911	3,282
Panel B: Second-order polynomial	(I)	(II)	(III)	(IV)	(V)	(VI)
Diff. in means	0.034	0.052**	0.059***	0.031***	0.077***	0.059***
p value	0.150	0.014	0.000	0.006	0.000	0.000
Bandwidth	1	2	3	4	5	6
#L	971	2,080	3,359	4,867	6,492	8,315
#R	764	1,428	1,974	2,481	2,911	3,282

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. P value indicates finite-sample p values obtained from the randomization test. Polynomial indicates the polynomial order used to estimate the regression, with the function allowed to differ on either side of the cutoff. This approach aligns with the idea of randomization inference for RDD under local randomization as proposed by Cattaneo et al. (2015) and Cattaneo et al. (2017) and implemented by Cattaneo et al. (2016).

References

Alicke, M.D., Davis, T.L., Pezzo, M.V., 1994. A posteriori adjustment of a priori decision criteria. Soc. Cogn. 12 (4), 281–308. doi:10.1521/soco.1994.12.4.281.

Allen, E.J., Dechow, P.M., Pope, D.G., Wu, G., 2017. Reference-dependent preferences: evidence from marathon runners. Manag. Sci. 63 (6), 1657–1672. doi:10.1287/mnsc.2015.2417.

Bachmann, K., 2018. Can advisors eliminate the outcome bias in judgements and outcome-based emotions? Rev. Behav. Financ. 10 (4), 336–352. doi:10.1108/RBF-11-2016-0072.

Baron, J., Hershey, J.C., 1988. Outcome bias in decision evaluation. J. Pers. Soc. Psychol. 54 (4), 569. doi:10.1037/0022-3514.54.4.569.

Beilock, S.L., Carr, T.H., MacMahon, C., Starkes, J.L., 2002. When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. J. Exp. Psychol.: Appl. 8 (1), 6. doi:10.1037/1076-898X.8.1.6.

Berger, J., Pope, D., 2011. Can losing lead to winning? Manag. Sci. 57 (5), 817–827. doi:10.1287/mnsc.1110.1328.

Billett, M.T., Qian, Y., 2008. Are overconfident CEOs born or made? Evidence of self-attribution bias from frequent acquirers. Manag. Sci. 54 (6), 1037–1051. doi:10.1287/mnsc.1070.0830.

Brownback, A., Kuhn, M.A., 2019. Understanding outcome bias. Games Econ. Behav. 117, 342–360. doi:10.1016/j.geb.2019.07.003.

Bühren, C., Krabel, S., 2019. Human performance after success and failure: evidence from the NBA. Int. J. Hum. Resour. Manag. 1–24.

Calonico, S., Cattaneo, M.D., Farrell, M.H., Titiunik, R., 2017. rdrobust: software for regression-discontinuity designs. Stata J. 17 (2), 372–404. doi:10.1177/1536867X1701700208.

Calonico, S., Cattaneo, M.D., Farrell, M.H., Titiunik, R., 2019. Regression discontinuity designs using covariates. Rev. Econ. Stat. 101 (3), 442–451. doi:10.1162/rest_a_00760.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014a. Robust data-driven inference in the regression-discontinuity design. Stata J. 14 (4), 909–946. doi:10.1177/1536867X1701700208.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014b. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82 (6), 2295–2326. doi:10.3982/ECTA11757.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2015. Randomization inference in the regression discontinuity design: an application to party advantages in the US Senate. J. Causal Inference 3 (1), 1–24. doi:10.1515/jci-2013-0010.

Cattaneo, M.D., Idrobo, N., Titiunik, R., 2019. A practical introduction to regression discontinuity designs: Foundations. Cambridge University Press.

Cattaneo, M.D., Titiunik, R., Vazquez-Bare, G., 2016. Inference in regression discontinuity designs under local randomization. Stata J. 16 (2), 331–367. doi:10.1177/1536867X1601600205.

Cattaneo, M.D., Titiunik, R., Vazquez-Bare, G., 2017. Comparing inference approaches for RD designs: a reexamination of the effect of head start on child mortality. J. Policy Anal. Manag. 36 (3), 643–681. doi:10.1002/pam.21985.

Cushman, F., Dreber, A., Wang, Y., Costa, J., 2009. Accidental outcomes guide punishment in a "trembling hand" game. PLoS One 4 (8), e6699. doi:10.1371/journal.pone.0006699.

Czaja, D., Röder, F., 2020. Self-attribution bias and overconfidence among nonprofessional traders. Q. Rev. Econ. Financ. 78, 186–198. doi:10.1016/j.qref.2020.02.003.

Emerson, G.B., Warme, W.J., Wolf, F.M., Heckman, J.D., Brand, R.A., Leopold, S.S., 2010. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. Arch. Intern. Med. 170 (21), 1934–1939. doi:10.1001/archinternmed.2010.406.

Federal Chancellery. (2022, May 26). *Cost of learning to drive*. https://www.ch.ch/en/vehicles-and-traffic/driving-licence/learning-to-drive/cost-of-learning-to-drive/.

Federal Statistical Office. (2021, July 16). Bruttoerwerbseinkommen pro Jahr der Erwerbstätigen nach Erwerbsstatus, Berufsgruppen ISCO 08, Beschäftigungsgrad und Geschlecht - Zentralwert (Median) in Franken. https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/loehne-erwerbseinkommen-arbeitskosten.assetdetail.17604449.html.

Federal Statistical Office. (2022, February 21). *Besitz von Fahrzeugen, Führerausweisen und ÖV-Abos*. https://www.bfs.admin.ch/bfs/de/home/statistiken/mobilitaet-verkehr/personenverkehr/verkehrsverhalten/besitz-fahrzeuge-fahrausweise.html.

Fitts, P.M., 1964. Perceptual-motor skill learning. Categories of Human Learning. Elsevier, pp. 243–285.

Fitts, P.M., 1965. Factors in complex skill training. In R. Glaser (Ed.), *Training Research and Education* (Vol. 1962, pp. 177–197). J. Wiley.

Fitts, P.M., Posner, M.I., 1967. Human Performance. Brooks/Cole.

Fleishman, E.A., 1965. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), Training Research and Education (pp. 137–175). J. Wiley.

Flepp, R., 2021. Uninformative performance signals and forced CEO turnover. Available at SSRN 3904056. Advance Online Publication. doi:10.2139/ssrn.3904056.

Flepp, R., Franck, E., 2021. The performance effects of wise and unwise managerial dismissals. Econ. Inq. 59 (1), 186–198. doi:10.1111/ecin.12924.

Gauriot, R., Page, L., 2019. Fooled by performance randomness: overrewarding luck. Rev. Econ. Stat. 101 (4), 658–666. doi:10.1162/rest_a_00783.

Gelman, A., Imbens, G., 2019. Why high-order polynomials should not be used in regression discontinuity designs. J. Bus. Econ. Stat. 37 (3), 447–456. doi:10.1080/07350015.2017.1366909.

Gillenkirch, R.M., Velthuis, L., 2018. Subjective evaluations of risk taking decisions: experimental evidence on outcome biases and their consequences. Available at SSRN 3232703. Advance online publication. doi:10.2139/ssrn.3232703.

Gino, F., Shu, L.L., Bazerman, M.H., 2010. Nameless+ harmless= blameless: when seemingly irrelevant factors influence judgment of (un) ethical behavior. Organ. Behav. Hum. Decis. Process. 111 (2), 93–101. doi:10.1016/j.obhdp.2009.11.001.

Gurdal, M.Y., Miller, J.B., Rustichini, A., 2013. Why blame? J. Political Econ. 121 (6), 1205–1247. doi:10.1086/674409.

Hershey, J.C., Baron, J., 1992. Judgment by outcomes: when is it justified? Organ. Behav. Hum. Decis. Process. 53 (1), 89–93. doi:10.1016/0749-5978(92)90056-D.

Hilary, G., Hsu, C., 2011. Endogenous overconfidence in managerial forecasts. J. Account. Econ. 51 (3), 300–313. doi:10.1016/j.jacceco.2011.01.002.

Hilary, G., Menzly, L., 2006. Does past success lead analysts to become overconfident? Manag. Sci. 52 (4), 489–500. doi:10.1287/mnsc.1050.0485.

Jones, S.K., Yurak, T.J., Frisch, D., 1997. The effect of outcome information on the evaluation and recall of individuals' own decisions. Organ. Behav. Hum. Decis. Process. 71 (1), 95–120. doi:10.1006/obhd.1997.2714.

Kausel, E.E., Ventura, S., Rodríguez, A., 2019. Outcome bias in subjective ratings of performance: evidence from the (football) field. J. Econ. Psychol. 75, 102132. doi:10.1016/j.joep.2018.12.006.

Kieren, P., Weber, M., 2020. *Expectation formation under uninformative signals.* doi:10.2139/ssrn.3971733.

Klein Teeselink, Bouke, van den Assem, Martijn J., van Doler, Dennie, 2022. Does losing lead to winning? An empirical analysis for four different sports. Management Science doi:10.1287/mnsc.2022.4372.

Knowlton, B., Siegel, A., Moody, T., 2017. Procedural learning in humans. In doi:10.1016/B978-0-12-809324-5.21085-7.

König-Kersting, C., Pollmann, M., Potters, J., Trautmann, S.T., 2021. Good decision vs. good results: outcome bias in the evaluation of financial agents. Theory Decis. 90 (1), 31–61. doi:10.1007/s11238-020-09773-1.

Lefgren, L., Platt, B., Price, J., 2015. Sticking with what (barely) worked: a test of outcome bias. Manag. Sci. 61 (5), 1121–1136. doi:10.1287/mnsc.2014.1966.

Lewin, I., 1982. Driver training: a perceptual-motor skill approach. Ergonomics 25 (10), 917–924. doi:10.1080/00140138208925051.

Lipshitz, R., 1989. Either a medal or a corporal": the effects of success and failure on the evaluation of decision making and decision makers. Organ. Behav. Hum. Decis. Process. 44 (3), 380–395. doi:10.1016/0749-5978(89)90015-0.

Marshall, G.W., Mowen, J.C., 1993. An experimental investigation of the outcome bias in salesperson performance evaluations. J. Personal Sell. Sales Manag. 13 (3), 31–47. doi:10.1080/08853134.1993.10753956.

Murad, Z., Starmer, C., 2021. Confidence snowballing and relative performance feedback. J. Econ. Behav. Organ. 190, 550–572. doi:10.1016/j.jebo.2021.08.006.
Peecher, M.E., Piercey, M.D., 2008. Judging audit quality in light of adverse outcomes: evidence of outcome bias and reverse outcome bias. Contempor. Account. Res. 25 (1), 243–274. doi:10.1506/car.25.1.10.
Pope, D., Simonsohn, U., 2011. Round numbers as goals: evidence from baseball, SAT takers, and the lab. Psychol. Sci. 22 (1), 71–79. doi:10.1177/0956797610391098.
Proud, S., 2015. Resits in higher education: merely a bar to jump over, or do they give a pedagogical 'leg up'? Assess. Eval. Higher Educ. 40 (5), 681–697. doi:10.1080/02602938.2014.947241.
Puetz, A., Ruenzi, S., 2011. Overconfidence among professional investors: evidence from mutual fund managers. J. Bus. Financ. Account. 38 (5–6), 684–712. doi:10.1111/j.1468-5957.2010.02237.x.
Ratner, R.K., Herbst, K.C., 2005. When good decisions have bad outcomes: the impact of affect on switching behavior. Organ. Behav. Hum. Decis. Process. 96 (1), 23–37. doi:10.1016/j.obhdp.2004.09.003.
Rubin, J., Sheremeta, R., 2016. Principal–agent settings with random shocks. Manag. Sci. 62 (4), 985–999. doi:10.1287/mnsc.2015.2177.
Sanchez, D.J., Reber, P.J., 2013. Explicit pre-training instruction does not improve implicit perceptual-motor sequence learning. Cognition 126 (3), 341–351. doi:10.1016/j.cognition.2012.11.006.
Sezer, O., Zhang, T., Gino, F., Bazerman, M.H., 2016. Overcoming the outcome bias: making intentions matter. Organ. Behav. Hum. Decis. Process. 137, 13–26. doi:10.1016/j.obhdp.2016.07.001.
Soll, J.B., Milkman, K.L., Payne, J.W., 2014. *A User's Guide to Debiasing.*
Tinsley, C.H., Dillon, R.L., Cronin, M.A., 2012. How near-miss events amplify or attenuate risky decision making. Manag. Sci. 58 (9), 1596–1613. doi:10.1287/mnsc.1120.1517.