A double-edged sword: How and why resetting performance metrics affects motivation and performance

Hengchen Dai*

University of California, Los Angeles, United States

ARTICLE INFO

ABSTRACT

Inside and outside of workplaces, individuals' performance on a metric (e.g., sales) is often decoupled from past performance (rather than being tracked as a continuation of past performance). How do people respond to such performance resets, a type of fresh start on performance records, particularly when resets are not anticipated? Three laboratory experiments and one field study analyzing 40 years of data from professional baseball players demonstrate their impact. Specifically, unanticipated resets increase self-efficacy and thus boost motivation and future performance when they follow weak performance. However, such resets decrease self-efficacy and thus harm motivation and future performance when they follow strong performance. By identifying the conditions that determine whether performance resets improve or harm motivation, and highlighting the role of self-efficacy, this paper provides novel insights into how different ways of tracking performance influence motivation, as well as how fresh starts change behavior.

1. Introduction

In the era of big data and the quantified self, where individual performances can be easily tracked, performance feedback has become more ubiquitous than ever (Benartzi & Lehrer, 2015; Wilson, 2013). Companies have begun to revolutionize performance management by making use of advances such as employee monitoring software, workplace tracking devices, and feedback-tracking apps, as well as adjusting methods of providing performance feedback (Cappelli & Tavis, 2016). Technology enables managers to deliver and refresh feedback at a higher and more flexible frequency without being restricted by traditional monthly, quarterly, and yearly cycles (Benartzi & Lehrer, 2015). Thus, it is important for managers and individual decision-makers to be aware of the impact that performance tracking may have on motivation.

In this paper, I study how *performance resets*—moments when individuals' performance on a metric is decoupled from past performance—affect motivation and future performance. In a survey of 572 full-time employees recruited through Qualtrics and Amazon's Mechanical Turk (average work experiences at the current organization = 8.59 years), 40.03% of employees from a variety of occupations (e.g., business and financial operations, healthcare support, sales, office and administrative support, management, and education) provided a specific example of a performance reset in the workplace that fit my

definition, describing incidents at work in which their performance metrics were tracked from a new starting point independent of their past performance (see Appendix A for details about the survey method and results). Common examples included situations in which performance on a specific metric (e.g., sales tallies, billable hours, customer satisfaction) was reset to zero at the beginning of a calendar cycle (e.g., a month, a year), supervisor evaluations were started anew when a new manager evaluated employees independently from their previous manager, and performance records were tracked from a clean slate when evaluation standards changed (e.g., due to reorganizations, changes in company ownership, mergers and acquisitions). Beyond the workplace, performance resets may also occur in other contexts. For example, a student's grade point average (GPA) may be tracked independent of past GPA after transferring to a new school, and a user of a fitness device such as Fitbit may have her exercise statistics reset to zero each week.

My definition of performance reset focuses on how performance at a given point is *tracked*, but does not implicate how performance is *appraised* and *rewarded*. The extent to which performance prior to the reset applies toward overall performance appraisals and final rewards varies. For example, when a new manager joins in the middle of the year and evaluates employees independently of the previous manager, employees' supervisor evaluations experience a reset by my definition; however, the new manager may choose to discard, partially consider, or

fully incorporate the previous manager's evaluations for annual performance appraisals and bonuses at the end of the year. Regardless of the new manager's decision about the relevance of the previous manager's supervisor evaluations, this scenario reflects the essence of performance resets in that the new manager's evaluations are decoupled from evaluations given by the previous manager. To focus on this essence of performance resets, my research examines whether the mere act of *tracking* performance from a new starting point, without altering the financial implications of past performance, can influence motivation.

On one hand, there are reasons to believe that merely tracking performance from a new starting point can boost motivation. Recent research on the "fresh start effect" (Dai, Milkman, & Riis, 2014, 2015) has demonstrated that people are motivated to tackle their self-improvement goals following temporal landmarks—dates that signal the beginning of a new calendar period (e.g., the start of the week/month/year, holidays) or significant personal events (e.g., birthdays, marriages; Shum, 1998). Though performance resets concern the tracking of performance rather than time and may not necessarily co-occur with temporal landmarks,[1] performance resets and temporal landmarks share one important similarity: both are a type of fresh start that separate the past and the present. In this sense, prior work about temporal landmarks suggests that people feel more motivated when their current performance is tracked separately from their past performance, implying that a performance reset induces a boost in motivation.

On the other hand, this prediction may be an oversimplification if varying levels of past performance actually lead people to respond differently to performance resets. Studies on the fresh start effect (Dai et al., 2014, 2015) have focused on people who want to improve specific aspects of themselves (e.g., to eat more healthily, exercise more) often by explicitly recruiting people who have failed to achieve their goals (e.g., Studies 2 and 4 in Dai, Milkman, & Riis, 2015). This raises the question of whether the motivating power of having a fresh start documented in prior work only applies when people had past failures. In the context of the current research, will tracking performance from a clean slate still motivate people who recently performed well? Anecdotally, salespeople often experience the "first-of-the-month blues", or feelings of discouragement when their excellent sales records are reset to zero at the start of a new month (O'Brien, 2015). As a sales coaching firm describes, "You were the Top Sales Rep last month? That's great. Unfortunately, the month starts over…The sales-goal reset… undermines good work. For some, it can even lead to performance-threatening burnout" (O'Brien, 2015). In other words, for salespeople with a strong sales record in the previous month, starting over with a clean sales record may feel like a daunting task, even when rewards for past performance have already been acquired. Studying how people with varying levels of past performance adjust effort after a reset will shed light on when and for whom a performance reset can be motivating vs. demotivating.

In this research, I propose that performance resets can either boost or harm motivation and future performance, depending on past performance: performance resets (relative to no resets) increase self-efficacy and thus boost motivation and future performance following weak past performance, but reduce self-efficacy and thus harm motivation and future performance following strong past performance. In the following sections, I develop my hypotheses and review my theoretical contributions before presenting three laboratory experiments and a field study that explore how and why performance resets affect motivation and future performance.

---

[1] Performance resets may happen without temporal landmarks (e.g., a manager may decide to use new attendance tracking software in the middle of the year, as opposed to at the beginning of a fiscal cycle), and performance resets may not accompany temporal landmarks (e.g., for some salespeople, sale totals are not reset each month).

1.1. Performance resets and mental accounting

Environmental cues and physical acts that decouple past and future events, even in a symbolic way, can close the "mental accounts" associated with past events and open new "mental accounts" (Imas, 2016; Li, Wei, & Soman, 2010; Read, Loewenstein, & Rabin, 1999; Soster, Monga, & Bearden, 2010; Tu & Soman, 2014). For example, transitions into new calendar periods cause people to write off temporal costs of an item (i.e., the time spent acquiring it) and disassociate the costs with the future benefits derived from the item (Soster et al., 2010). Highlighting a temporal landmark between now and a task deadline puts the deadline and the present moment into two different categories and causes people to put off working on the task (Tu & Soman, 2014). Additionally, physically enclosing a written recollection about a regretted past decision into an envelope provides a sense of closure and alleviates people of the negative emotions induced by the unpleasant past experiences (Li et al., 2010); actions that realize prior losses (e.g., cashing out losses at the gambling table, selling losing stocks) close mental accounts associated with prior losses and reset the reference point against which people evaluate subsequent prospects, reducing the likelihood that people try to recoup their losses (Imas, 2016). Altogether, this literature suggests that by decoupling future performance from past performance, performance resets cause people to view past and future performance outcomes as in two separate mental accounts.

Past research further suggests that people tend to contrast objects in separate mental categories and assimilate objects in the same category (Mishra & Mishra, 2010; Peetz & Wilson, 2013; Schwarz & Bless, 2007). For example, a distance between two points depicted in the same spatial group is perceived as smaller than the objectively identical distance between two points depicted in different spatial groups (Huttenlocher, Hedges, & Duncan, 1991; Maddox, Rapp, Brion, & Taylor, 2008; Tversky, 1992). Also, by creating boundaries between temporal periods, past temporal landmarks make people feel less similar to their former selves from the pre-landmark era (Dai et al., 2015). Similarly, upcoming temporal landmarks lead people to contrast their current selves with their future selves associated with the post-landmark era (Peetz & Wilson, 2013). This literature suggests that as performance resets separate past and future performance into two mental accounts, people tend to contrast (as opposed to assimilate) their future outcomes with their past outcomes. I expect this tendency to have implications for self-efficacy, which I theorize about in the next section.

1.2. Performance resets, past performance, and self-efficacy

Self-efficacy is a person's belief in her ability to perform well in a given task (Bandura, 1997). One strong determinant of self-efficacy is an individual's past performance: Past successes indicate that one is capable of harnessing the resources required to perform well and elicit high self-efficacy, whereas past failures cast doubt on future accomplishments and decrease self-efficacy (Bandura, 1991, 1997; Locke, Frederick, Lee, & Bobko, 1984).

However, when people believe that their past performance does not accurately indicate their current capabilities, the effect of past performance on their self-efficacy is diminished. For example, past failures dampen self-efficacy to a lesser extent when attributed to external (vs. internal) factors, and similarly, past successes boost self-efficacy to a lesser extent when attributed to external (vs. internal) factors (Silver, Mitchell, & Gist, 1995; Tolli & Schmidt, 2008). External attributions reduce people's tendency to view past performance as reflective of their current capabilities, resulting in a diminished effect of past performance on self-efficacy. Also, the more distance people feel from their past outcomes, the weaker the extent to which those past outcomes shape their beliefs about their current and future prospects. For instance, Libby and Eibach (2002) show that when dieters use a third-person (vs. a first-person) perspective to recall past episodes of overindulgent eating, they become more optimistic about their ability to refrain from

overeating during their next Thanksgiving dinner. In other words, the third-person perspective distances people from their undesirable past actions and makes them feel less tarnished by their past failures. In a similar vein, it is more difficult for people to view their current selves positively when past successes feel far away from the present (Broemer, Grabowski, Gebauer, Ermel, & Diehl, 2008; Ross & Wilson, 2002). Furthermore, evaluations of the past are less likely to influence people's view of their current state when they are primed with a contrast mindset than when they are primed with a similarity mindset (Hanko, Crusius, & Mussweiler, 2010).

Connecting research on past performance and self-efficacy with research on mental accounting, I theorize that past performance and performance resets jointly influence self-efficacy. By separating past performance into a different mental account from current performance, a reset may lead people to perceive their past performance as less indicative of their current capabilities. This stands in contrast to situations without a reset where future performance is continuously tracked from and naturally assimilated with past performance. As a result, following *weak* past performance, I predict that a reset will reduce people's tendency to base their self-efficacy on past failures and thus make them feel *more* confident than they would otherwise. Similarly, following *strong* past performance, I predict that a reset will reduce people's tendency to base their self-efficacy on their past successes and thus make them feel *less* confident than they would otherwise.

1.3. Performance resets, past performance, motivation, and future performance

Self-efficacy has been shown to be a predictor of motivation and performance. Compared to those with low self-efficacy, individuals with high self-efficacy select more challenging tasks (Bandura & Schunk, 1981), set more challenging goals (Brown, Jones, & Leigh, 2005; Tolli & Schmidt, 2008), exert more effort (Schunk & Hanson, 1985; Schunk, Hanson, & Cox, 1987), work more persistently when faced with difficulties (Multon, Brown, & Lent, 1991), and eventually achieve higher performance (Bandura & Locke, 2003; Brown et al., 2005; Stajkovic & Luthans, 1998). Self-efficacy also affects task selection: people tend to avoid tasks for which they have a low level of self-efficacy and favor those for which they have high self-efficacy (Bandura, 1997; Sullivan, O'Connor, & Burris, 2006).

Combining the positive relationship between self-efficacy, motivation, and future performance with the arguments in the previous section, I propose that relative to situations without resets, performance resets can protect self-efficacy and thus boost motivation and future performance among individuals with weak past performance, but reduce self-efficacy and thus harm motivation and future performance when individuals have strong past performance. Formally, I hypothesize:

H1a-H1b: Past performance will moderate the effects of performance resets on motivation (H1a) and future performance (H1b). Specifically, resets will increase motivation and future performance following weak past performance but decrease motivation and future performance following strong past performance.

H2: Past performance will moderate the effects of performance resets on self-efficacy. Specifically, resets will increase self-efficacy following weak past performance but decrease self-efficacy following strong past performance.

H3a-H3b: Self-efficacy will mediate the relationship between performance resets, past performance, and motivation (H3a) as well as future performance (H3b).

It is worth noting that some performance resets are anticipated (e.g., an individual realizes resets occur monthly), while others are not (e.g., an organizational change produces a reset without warning). I expect that my predictions hold for both types of performance resets, but it is possible that anticipating a performance reset could also change an individual's effort *before* a reset occurs. In this paper, I focus on testing

my hypotheses in settings where individuals do not anticipate a performance reset; thus, I can hold information and motivation constant across experimental conditions before a reset arrives and cleanly measure the moderating effect of past performance on the response to a reset. Future research that extends this work to study anticipated performance resets would be valuable.

This research makes several theoretical contributions, particularly vis-à-vis past research on temporal landmarks. First, by studying the effects of performance resets—which are prevalent inside and outside of workplaces—this research generates new insights into how different ways of tracking performance influence motivation. I examine the psychological effects of mere separation of past and future performance, holding the presence of temporal landmarks constant across conditions. Second, extending recent research on positive behavior changes associated with moments inducing a fresh start (Ayal & Gino, 2011; Dai et al., 2014, 2015; Hennecke & Converse, 2017), this paper investigates both the positive and negative aspects of a novel type of fresh start (i.e., performance resets). Third, I develop a theory that predicts when performance resets will boost or decrease motivation. My theory extends the identity-based theory about psychological distance between temporal selves that has been used to explain the effects of temporal landmarks (Dai et al., 2014, 2015; Peetz & Wilson, 2013). Taken together, this research uncovers previously untapped antecedents of motivation and provides a more complete view of how and why fresh starts affect behavior.

2. Overview of studies

I present three laboratory experiments and one field study that investigate the effects of performance resets, with a particular focus on unanticipated resets. All of my studies compare performance resets to a no-reset control condition where present performance is tracked as a continuation of past performance, allowing me to eliminate regression to the mean as an alternative explanation for my findings. In Studies 1–3, I introduced a performance reset by merely tracking future performance separately from past performance without altering the financial implications of past performance, and individuals were unaware of the reset in advance. In Study 1, I measured my theoretical moderator—the level of past performance. In Studies 2 and 3, I additionally manipulated past performance, tested my full model where self-efficacy is the underlying mechanism, and addressed alternative explanations. Finally, Study 4 examined a high-stakes field setting, relying on performance data about professional baseball players who were traded in the middle of a regular baseball season over the past 40 years. In this context where players often do not know in advance whether and when they will be traded, I assessed how baseball players' performance after a trade varies as a function of (a) whether their trade triggers a performance reset and (b) their performance prior to the trade. Altogether, the four studies demonstrate that past performance moderates the effects of unanticipated performance resets on motivation and future performance by altering people's self-efficacy. In all experiments, I determined the sample sizes in advance of data collection. Study materials are available at https://osf.io/e6cyb/, and details about recruitment procedures are reported in Appendix B.

3. Study 1: Word search task

In Study 1, I manipulated the presence of an unanticipated performance reset and measured participants' performance before and after the reset. I investigated how measured past performance moderated the effect of performance resets on future performance (H1b).

3.1. Method

3.1.1. Participants

I recruited 202 participants (41.71% female, M_{age} = 34.49, three

unspecified) from Amazon's Mechanical Turk (MTurk) to take part in a 15-min study.

3.1.2. Procedure

Participants were informed that the study involved 10 one-minute Boggle games and that they would be paid $0.02 for every word they correctly generated during the study, in addition to a base pay. For each game, participants were presented with a 3x3 grid of nine letters and searched for words that (a) consisted of three or more letters adjoining vertically, horizontally, or diagonally and (b) did not reuse a letter.

After passing comprehension check questions about the game rules and playing a practice game, participants proceeded to the 10 actual games. After each of the first five games, all participants received performance feedback in the same manner. Specifically, a bar graph that marked Games 1–10 on the X-axis depicted the number of correct words participants generated each game. Games that were yet to come had a score of zero. Participants were also informed of their average score (i.e., the average number of correct words generated per game) across all of the games they had already completed. There was one horizontal line across the bar graph indicating the average number of correct words per game earned by 21 pretest participants in the same sample (see Fig. 1, Panel A). This information about peers' performance served as a benchmark for participants to evaluate their own performance.

I randomly assigned participants to either the *control* or *performance-reset* condition after five games. Participants in the control condition read, "You have finished five games. Game 6 will begin in one minute. We will continue tracking your average score when the games resume." In the next five games (i.e., Game 6–10), participants in the control condition received feedback in the same manner they had in the first five games, and their running average score was tracked from Game 1 up to Game 10.

In the performance-reset condition, participants read, "You have finished the first round of five games. Your average score on Round 1 has been saved. A new round will begin in one minute. We will track your average score on Round 2 from zero, a new starting point." Participants were then presented with a new performance interface designed to only track scores for the next five games (see Fig. 1, Panel B). After each game, the bar graph updated participants' scores and tracked their running average score in Round 2 from the first game of Round 2 (which was essentially Game 6). Notably, participants' average score across the first five games was still displayed at the top of the feedback interface. This intended to address the concern that participants may forget about their performance in the first five games or mistakenly think their past performance did not count towards their final payoff.

In this and all subsequent experiments, I conducted manipulation checks and confirmed that my manipulations were effective (see Appendix B for measures and results). Also, in all experiments, participants reported their age and gender at the end of the study.

3.2. Results and discussion

I measured *past performance* and *future performance* as the number of correct words participants generated in the first five games ($M = 35.19$, $SD = 15.47$) and last five games ($M = 37.75$, $SD = 15.00$), respectively. In an ordinary least squares (OLS) regression, I predicted future performance using a dummy indicator for the performance-reset (versus control) condition, a mean-centered continuous measure of past performance, and the interaction between the performance-reset indicator and (mean-centered) past performance. The negative and significant interaction term ($B = -0.20$, $SE = 0.08$, $p = 0.009$, 95% CI = [-0.35, -0.05], $\eta_p^2 = 0.03$) suggests that the performance-reset treatment influenced future performance differently depending on past performance. I conducted a simple slopes analysis following the steps outlined by Aiken and West (1991). When past performance was one standard deviation below the mean, a performance reset was estimated to marginally significantly increase future performance (a gain of 2.65 words relative to the control; $p = 0.10$, 95% CI = [-0.55, 5.85]); however, a reset was estimated to significantly decrease future performance when past performance was one standard deviation above the mean (a loss of 3.58 words relative to the control; $p = 0.03$, 95% CI = [0.31, 6.84]).

In Study 1, the reset manipulation simply tracked participants' average score in the last five games from a new starting point but still counted their past performance toward their final pay. I took several measures to reduce the possibility that participants in the reset condition incorrectly believed that their past performance was irrelevant to their final pay. First, participants were told clearly at the beginning of the study that they would be paid for every word they generated during the study. Second, as explained in the Method section, my reset manipulation did not indicate any change in the incentive structure and assured participants that their scores in the first five trials were saved. Third, participants in the reset condition were presented with their average scores on the first five games while working on the last five games. With this framing-based reset manipulation, Study 1 showed that past performance moderated the effects of a performance reset on future performance, supporting H1b. In Studies 2 and 3, I examined both motivation and future performance, measured self-efficacy, and tested my full mediated moderation model. Additionally, the next two studies addressed the concern that differing past performance in Study 1 was endogenous by manipulating past performance (Study 2) or individuals' perceptions of their past performance (Study 3).

Panel A: This interface template was used in the first five games for all participants and in the last five games for participants in the control condition.

Panel B: This interface template was used in the last five games for participants in the performance-reset condition.
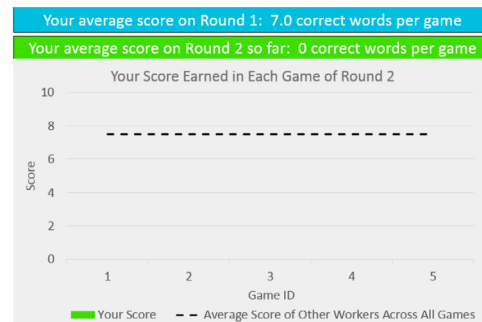

Fig. 1. Templates of performance feedback interfaces used in Study 1. Panel A: This interface template was used in the first five games for all participants and in the last five games for participants in the control condition. Panel B: This interface template was used in the last five games for participants in the performance-reset condition.

4. Study 2: Personal goal pursuit and psychological mechanisms

In Study 2, I experimentally manipulated both the presence of an unanticipated performance reset and strength of past performance. Inspired by the abundant apps allowing people to track their goal progress and habit formation, I tested the effects of resets in a context where people track their performance statistics about personal goals using an app. In this context, people do not receive financial incentives for achieving personal goals, nor are there concerns about public knowledge of performance or reputation. In addition to testing my proposed mechanism of self-efficacy, I also explored alternative explanations suggested by past research.

4.1. Method

4.1.1. Participants

On October 1, 2017, I recruited 361 participants (59.28% female, M_{age} = 36.80) from MTurk who wanted to develop a habit of performing a specific positive activity every day but had not yet completed that activity on that day at the time of taking my survey.

4.1.2. Procedure

Participants first indicated a specific positive activity they wanted to form a habit of doing every day by either selecting one from a list of seven activities—which were adapted from common activities on the popular habit-tracking app Strides (e.g., walking at least 5,000 steps, sleeping by 11 pm)—or briefly describing their own activity. Then, they were asked to imagine that they had begun to work on this habit three weeks ago on September 10, 2017, and had been using an app to keep track of their progress. The app allowed people to report whether they did the focal activity every day and depicted progress with the aid of a calendar. Participants received a green dot for each day they successfully completed the focal activity, and a red dot for each day they did not. Consecutive green dots formed a streak, which would be broken by failing to complete the focal activity on a successive day. The app also displayed people's current streak (in days). The design of the app was based on two popular habit-tracking apps in the marketplace, Strides and Streaks.

Participants were presented with five progress reports on the app and were asked to imagine that they received these progress reports over the course of three weeks from September 10 to September 30, 2017. The first three progress reports were identical among participants. Specifically, the first report showed that participants started working on their habit on September 10. It displayed a calendar of September 2017, with September 10 marked by a yellow dot. The second report showed that the focal activity was completed on 50% of the days in the first week after participants started working on their habit, and the third report showed that the focal activity was completed 50% of the time in the first two weeks. I then orthogonally manipulated participants' progress on the focal activity and whether they experienced a performance reset.

4.1.2.1. Weak vs. strong past performance. I manipulated feedback in the fourth and fifth progress reports, which reflected participants' hypothetical progress as of September 27 and September 30, respectively. In the weak-past-performance condition, the two progress reports indicated that participants did *not* consistently perform the focal activity from September 23 to September 30. In the strong-past-performance condition, the two progress reports indicated that participants consistently performed the focal activity from September 23 to September 30.

4.1.2.2. Performance reset vs. control. After participants viewed the aforementioned five progress reports, I manipulated the presence of a performance reset by varying the format of progress report for October 1 (the day the experiment was conducted). Participants in the control condition were presented with a screenshot of the app which displayed their progress in September 2017 above a clean calendar of October 2017. They were told that their statistics for the habit would be continuously tracked from September to October. Participants in the performance-reset condition were presented with a screenshot of the app showing only a clean calendar of October 2017. They were told that their statistics for the habit would be tracked from a new starting point in October, and that their progress in September could be accessed by swiping to a different page in the app. My reset manipulation was inspired by the design of popular habit-tracking apps: Streaks' calendar-view report only features users' progress in the current month; Strides resets users' current streak on the first day of a new month; and both apps allow users to view their progress in previous months by swiping to the right.

Next, I asked participants a few questions about their experiences. I assessed *motivation* by asking participants how motivated they would be to perform the focal activity today (1 = Not at all, 9 = Extremely motivated). Participants were then asked to explain in a few sentences why they would or would not be motivated to do the activity today. I measured *self-efficacy* by assessing participants' confidence in consistently performing the focal activity going forward. Specifically, I averaged their responses to two questions (r = 0.87; Bandura, 2006; Sullivan et al., 2006; Schmidt & DeShon, 2010): "How confident are you that you can do this activity daily from today on?" (1 = Not confident at all, 9 = Very confident) and "How likely do you think you are to do this activity daily from today on?" (1 = Not likely at all, 9 = Very likely).[2]

I also included measures of other potential explanations. First, past research has shown that people desire to improve sequences of outcomes (Loewenstein & Prelec, 1993) and are motivated by symbols of progress (Kivetz, Urminsky, & Zheng, 2006; Li & Soman, 2010; Nunes & Drèze, 2006). In Study 2, participants in the strong-past-performance condition imagined making steadily good progress and having formed a good streak of performance. Among them, those in the control condition may have been motivated to engage in the focal activity today in order to maintain their progress and their streak, whereas those in the reset condition may have lost that motivation because the reset restarted their streak. Though this cannot explain why a reset would increase motivation among participants in the weak-past-performance condition, I explored the potential role of this account. I assessed the *perceived importance* of doing the focal activity today using two

---

[2] In Bandura's (2006) guide for constructing self-efficacy measures, exemplar measures asked participants to rate their degree of confidence in doing a given task well or achieving a specific performance level on a 100-point scale (0 = Cannot do at all, 100 = Highly certain can do). Researchers have used variations of this scale to measure the strength of self-efficacy. For example, Schmidt and DeShon (2010) studied the moderating effects of performance ambiguity on the relationship between self-efficacy and performance, and asked participants to rate their confidence to reach a performance level on a 10-point scale. Sullivan et al. (2006) studied the relationship between self-efficacy and uses of negotiation tactics, and asked participants to rate their confidence that they could use a negotiation tactic successfully on a 100-point scale (0 = No confidence; 100 = Full confidence). The first item of my *self-efficacy* measure captured participants' confidence in performing the focal activity well and was consistent with past research (Bandura, 2006; Sullivan et al., 2006; Schmidt & DeShon, 2010). The second item was imperfect: Though it was intended to capture participants' subjective probability of obtaining good performance (as in Seo & Ilies, 2009), its wording may also capture participants' motivation to do the focal activity daily. Notably, my results are robust if I only use participant's response to the first item. Also, I asked two independent coders, blind to my hypotheses and experimental condition, to rate participants' open-ended responses about why they would or would not be motivated to do the focal activity today. The coders rated the extent to which each response signaled that the participant was confident in her ability to consistently do the focal activity from today onward (1 = Not confident at all, 9 = Very confident; r = 0.78). My results are robust if I use coders' ratings to measure self-efficacy.

questions ($r = 0.87$; Huang, Jin, & Zhang, 2017): "How important is it for you to do this activity today to build a streak?" (1 = Not important at all, 9 = Extremely important) and "How much value do you see in doing this activity today to build a streak?" (1 = No value at all, 9 = Extremely high value).

Second, past research suggests that physically sealing negative past events can induce psychological closure and alleviate negative emotions about the past events (Li et al., 2010). It is possible that a reset creates psychological closure for undesirable past performance and enhances motivation by reducing negative affect. This account cannot explain why a reset dampens motivation following excellent past performance. Nonetheless, I explored the potential role of this account by asking participants to report their *negative affect* (upset, nervous, and anxious; α = 0.85; Watson, Clark, & Tellegan, 1988; Li et al., 2010) on a nine-point scale (1 = Very slightly, 9 = Very much).[3] I also measured *positive affect* (inspired, excited, enthusiastic; α = 0.96; Watson et al., 1988) on the same scale to more comprehensively capture changes in affective states.

Finally, I tested the identity-based theory that prior work has proposed to explain the effects of temporal landmarks—that is, the theory that temporal landmarks disconnect temporal selves (Dai et al., 2014, 2015; Peetz & Wilson, 2013). I assessed *psychological dissociation* between past and current selves using two questions ($r = 0.81$; Dai et al., 2015): "How distant are you feeling now from your past self in September 2017?" (1 = Extremely close, 9 = Extremely far away) and "How different are you feeling now from your past self in September 2017?" (1 = Completely the same, 9 = Completely different).

To ensure that my results were not sensitive to the order in which motivation and mechanism measures were collected, I manipulated whether participants rated and explained their motivation before or after they responded to the measures of potential mechanisms. I found no significant effects of order and report results below by collapsing the two orders.

4.2. Results

4.2.1. Motivation

Consistent with H1a, a two-way ANOVA on motivation revealed a significant interaction between the manipulation of past performance and the performance-reset manipulation, $F(1, 357) = 46.76$, $p < 0.0001$, $\eta_p^2 = 0.12$. The planned contrast analysis revealed that when participants imagined weak past performance, a performance reset made them feel significantly more motivated ($M = 7.21$, $SD = 1.72$) than the control condition ($M = 6.06$, $SD = 2.52$), $F(1, 357) = 14.57$, $p = 0.0002$, 95% $CI_{difference} = [0.52, 1.79]$, $d = 0.54$. However, when participants imagined strong past performance, a performance reset made them feel significantly less motivated ($M = 6.16$, $SD = 2.37$) than the control condition ($M = 7.92$, $SD = 1.21$), $F(1, 357) = 34.36$, $p < 0.0001$, 95% $CI_{difference} = [1.21, 2.31]$, $d = 0.93$. Fig. 2 Panel A depicts the results.

4.2.2. Self-efficacy

Consistent with H2, a two-way ANOVA on self-efficacy revealed a significant interaction between past performance and performance reset manipulations, $F(1, 357) = 28.28$, $p < 0.0001$, $\eta_p^2 = 0.07$. The

---

[3] The literature on psychological closure usually refers to negative affect about the negative event that is closed (Li et al., 2010). Thus, a measure of negative affect that better matches this literature should have asked participants how they were feeling about their past performance (e.g., on a scale ranging from very poor to very good). However, such a question would duplicate the question that I used as the manipulation check of past performance (Appendix B). Therefore, I asked participants about their overall negative affect in the moment, which I expected to be influenced by their negative affect towards their past performance. This measure was limited in terms of not closely capturing the direct effect of psychological closure on negative affect.

planned contrast analysis showed that when participants imagined weak past performance, a performance reset increased self-efficacy ($M = 5.63$, $SD = 1.84$) relative to the control condition ($M = 4.35$, $SD = 2.12$), $F(1, 357) = 22.05$, $p < 0.0001$, 95% $CI_{difference} = [0.69, 1.87]$, $d = 0.64$. However, when participants imagined strong past performance, a performance reset reduced self-efficacy ($M = 6.02$, $SD = 1.78$) relative to the control condition ($M = 6.78$, $SD = 1.49$), $F(1, 357) = 7.94$, $p = 0.005$, 95% $CI_{difference} = [0.28, 1.24]$, $d = 0.46$. Fig. 2 Panel B depicts the results.

4.2.3. Mediation by self-efficacy

Next, I tested whether self-efficacy mediated the interactive effect of a reset and past performance on motivation (H3a), using PROCESS model 8 (Hayes, 2013). A 5000-sample bootstrap analysis estimated an indirect effect of the interaction between reset and past performance manipulations on motivation via self-efficacy as −1.31 ($SE = 0.28$), and the 95% bias-corrected CI of the indirect effect did not include zero ($[−1.90, −0.79]$). Furthermore, when past performance was weak, the indirect effect of a reset on motivation via self-efficacy was positive and significant (indirect effect = 0.82; $SE = 0.21$; 95% CI = $[0.42, 1.24]$); when past performance was strong, the indirect effect of a reset on motivation via self-efficacy was negative and significant (indirect effect = −0.49; $SE = 0.17$; 95% CI = $[−0.82, −0.18]$). Thus, H3a was supported.
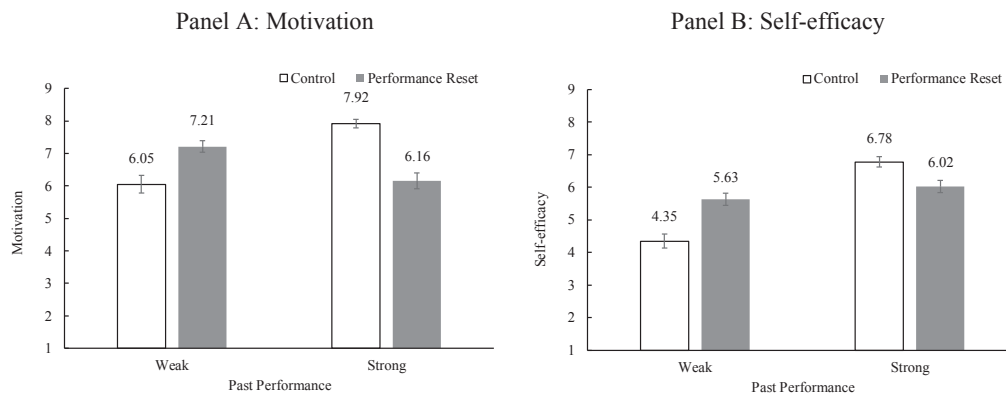
4.2.4. Alternative explanations

Two-way ANOVAs showed that reset and past performance manipulations significantly interacted to predict the perceived importance of doing the focal activity today ($F(1, 357) = 6.74$, $p = 0.01$, $\eta_p^2 = 0.02$), positive affect ($F(1, 357) = 17.21$, $p < 0.0001$, $\eta_p^2 = 0.05$), and negative affect ($F(1, 357) = 6.10$, $p = 0.01$, $\eta_p^2 = 0.02$). In terms of psychological dissociation from the past self, the interaction between reset and past performance manipulations was not significant, $F(1, 357) = 0.56$, $p = 0.45$, $\eta_p^2 = 0.002$. Notably, when I tested measures of alternative explanations along with self-efficacy as potential mediators using PROCESS model 8 (Hayes, 2013), self-efficacy remained a significant mediator of the interactive effect of reset and past performance on motivation (indirect effect = −0.42; $SE = 0.14$; 95% CI = $[−0.73, −0.17]$). The indirect effect of a reset on motivation via self-efficacy was 0.26 ($SE = 0.10$; 95% CI = $[0.10, 0.46]$) when past performance was weak and −0.16 ($SE = 0.07$; 95% CI = $[−0.31, −0.04]$) when past performance was strong. See Appendix C for how measures of alternative explanations differed by condition and the indirect effects via them.

4.3. Discussion

Study 2 examined a context where individuals pursued personal goals without concerns about financial incentives or reputation. To increase the realism and personal relevance of this scenario-based study, I asked participants to think about a goal they actually wanted to pursue (instead of asking them to imagine pursuing a hypothetical goal that they may or may not aspire to achieve). Across a variety of goals participants selected, I found that the effect of a reset on motivation to engage in a goal-related activity was moderated by past performance, whereby a reset increased individuals' motivation if they had not been consistently performing the activity, but decreased motivation if they had been consistently performing the activity. Additionally, Study 2 showed that the relationship between past performance, a reset, and motivation was driven by individuals' self-efficacy. By simultaneously testing multiple potential explanations, Study 2 highlighted the independent role played by self-efficacy. I replicated these results in two studies: one study used a different reset manipulation than Study 2, and another study used the same manipulation as Study 2 with minor modifications on design and measures (Appendix D).

Panel A: Motivation

Panel B: Self-efficacy

Note. Error bars represent the standard error of the mean.

Fig. 2. Motivation and self-efficacy as a function of performance reset and past performance manipulations (Study 2).

Notably, I conducted Study 2 on October 1—a calendar-defined temporal landmark that past research suggests could increase individuals' motivation to pursue personal goals (Dai et al., 2014)—and highlighted the beginning of a new month in both the reset and control conditions. This allowed me to test the effects of performance resets, while keeping the potential effects of temporal landmarks constant across conditions.

5. Study 3: Word unscrambling task and psychological mechanisms

Study 3 manipulated both the presence of an unanticipated reset and perceptions of past performance, tested my full model, and explored alternative explanations. Study 3 advances Studies 1 and 2 in three ways. First, Study 3 incentivized participants to perform well and involved both a behavioral measure of motivation and of future performance. Second, Study 3 eliminated the concern present in Study 1 that participants in the reset condition might incorrectly believe that their performance prior to the reset did not count towards their pay despite my efforts to reduce that possibility. In Study 3, I explicitly reminded participants before measuring their motivation that they would be paid based on their overall performance. Third, Study 3 used different manipulations of reset and past performance from previous studies, allowing me to establish the generalizability of the phenomenon across manipulations.

5.1. Method

5.1.1. Participants

I recruited 408 participants (42.36% female, four unspecified; M_{age} = 35.29, two unspecified) from MTurk to take part in a 20-min study.

5.1.2. Procedure

Participants were asked to perform a word task that involved rearranging letters to form English words. They were informed that each correct word was worth $0.15 and that one in five participants would be randomly chosen to receive the total dollar amount earned during the study, in addition to a base pay. For each trial, participants had 30 s to generate as many words as possible using all letters in a string of five or six letters. Participants were also informed that the program would determine whether their performance met researchers' expectations for each trial and provide feedback after every three trials. To increase the credibility of my performance feedback, I explained to participants that each trial had at least one and at most five correct solutions, and that researchers' expectations varied across trials based on the difficulty level of each trial.

After playing a practice game, participants played 12 actual trials.

After each trial, participants were told the actual number of correct words they generated. After every three trials, participants were additionally presented with a performance feedback interface that marked Trials 1–12 on the X-axis, which depicted a "thumbs up" badge for previous trials in which their performance met expectations, and a "thumbs down" badge for previous trials in which their performance did not meet expectations. The interface also displayed how long each participant's current streak was. All participants received the same feedback on the first three trials: success in Trial 1 and failure in Trials 2 and 3. Then I manipulated participants' perceptions of their performance as well as the presence of a performance reset.

5.1.2.1. Weak vs. strong past performance. In the weak-past-performance condition, participants received positive feedback only on two trials from Trial 4 to Trial 12. In the strong-past-performance condition, participants received positive feedback on all trials from Trial 4 to Trial 12. To increase credibility in my feedback, I chose letter strings based on pretests such that trials on which everyone received positive feedback had easy strings, trials on which everyone received negative feedback had difficult strings, and all other trials were of moderate difficulty.

5.1.2.2. Performance reset vs. control. After participants finished 12 trials, they were told that they would perform 12 more trials, and were randomly assigned to either the performance-reset or control condition. Participants in the control condition were told that as they proceeded with the word task, researchers would continue tracking and displaying their performance across all 24 trials. They were presented with a performance feedback interface that marked Trials 1–24 on the x-axis and displayed their progress on the first 12 trials (Fig. 3, Panels A and B), with either three trials receiving positive feedback (weak-past-performance condition) or nine trials receiving positive feedback (positive-past-performance condition). Participants in the performance-reset condition were told that as they proceeded with the word task, researchers would track and display their performance in the next 12 trials from a clean slate, independent of their performance from the first 12 trials. They were presented with a clean performance feedback interface that marked Trials 1–12 of Round 2 on the x-axis (Fig. 3, Panel C).

Right after my performance-reset manipulation, I collected measures of my proposed mechanism and alternative explanations. I assessed *self-efficacy* using one question (Bandura, 2006; Sullivan et al., 2006; Schmidt & DeShon, 2010): "How confident are you that you can build a streak of good performance in the next 12 trials?" (1 = Not confident at all, 9 = Very confident). I assessed the *perceived importance* of improving one's performance streak using two questions (r = 0.63; Huang et al., 2017): "How important is it for you to not break your current streak in the next trial?" and "How important is it for you to
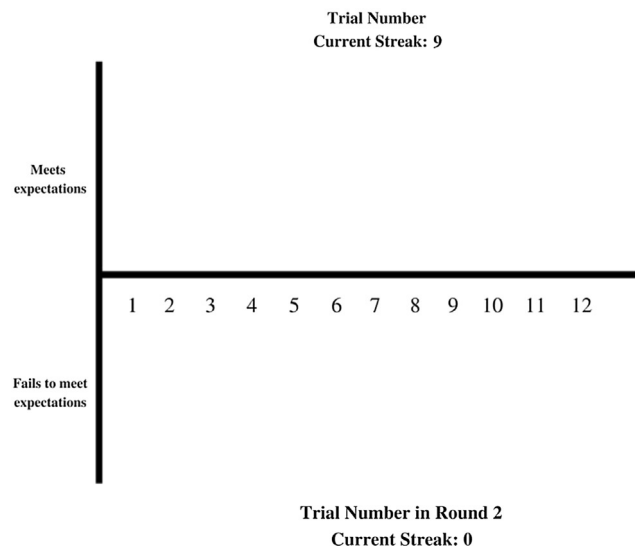
Panel A: After the first 12 trials, this interface was presented to participants in the control condition who were led to view their past performance as weak.

Panel B: After the first 12 trials, this interface was presented to participants in the control condition who were led to view their past performance as strong.
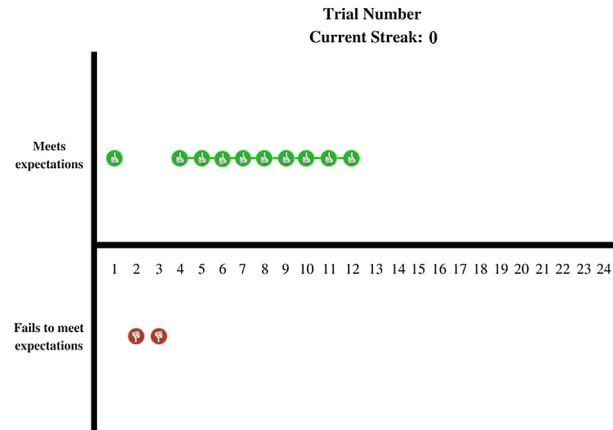
Panel C: After the first 12 trials, this interface was presented to participants in the performance-reset condition.

Fig. 3. Illustration of performance reset and past performance manipulations in Study 3. Panel A: After the first 12 trials, this interface was presented to participants in the control condition who were led to view their past performance as weak. Panel B: After the first 12 trials, this interface was presented to participants in the control condition who were led to view their past performance as strong. Panel C: After the first 12 trials, this interface was presented to participants in the performance-reset condition.

perform well in the next trial and extend your current streak?" (1 = Not important at all, 9 = Extremely important). Participants reported their *positive affect* (inspired, excited; $r = 0.91$) and *negative affect* (upset, discouraged; $r = 0.81$; Watson et al., 1988) on a nine-point scale (1 = Very slightly, 9 = Very much).

Next, participants chose to either complete another 12 trials of the word task or switch to an alternative task which involved answering 24 trivia questions in 12 trials. Participants learned how their bonus would be calculated if they were selected to receive a bonus: if they continued with the word task, they would be paid $0.15 for each correct word they generated across all 24 trials; if they switched to the trivia task,

they would be paid $0.15 for each correct word they generated in the first 12 trials and $0.10 for each correct trivia answer. For both tasks, participants knew that they had 30 s for each trial. My behavioral measure of *motivation* was whether participants chose to continue with the word task: persisting in the word task indicated greater motivation than quitting the word task.

Participants who chose to switch to the trivia task were directed to answer 12 trials of trivia questions. Participants who chose to proceed with the word task moved on to complete 12 trials of the word task. As in the first 12 trials, participants who continued with the word task learned the number of correct words they generated after each trial and

received feedback on whether their performance met researchers' expectations every three trials. Researchers' feedback was identical across participants such that nine of the 12 trials had positive feedback. Feedback interfaces either included or excluded participants' performance in the first 12 trials depending on whether they were in the control or performance-reset condition. Participants in the performance-reset condition could click a link beneath each feedback interface to review their performance in the first 12 trials. My measure of *future performance* was the number of correct words participants generated after the first 12 trials; this measure equaled zero for participants who switched to the trivia task and thus generated zero words after the first 12 trials.

5.2. Results

5.2.1. Motivation

In a logistic regression, I predicted participants' decision to stay in the focal task (the word task) using an indicator for the reset (vs. control) manipulation, an indicator for strong (vs. weak) past performance, and their interaction term. Consistent with H1a, past performance and reset manipulations significantly interacted to predict participants' motivation ($B = -1.57$, $SE = 0.42$, $p < 0.001$, 95% CI = $[-2.39, -0.76]$). Specifically, among participants who were led to view their past performance as weak, a reset significantly increased their likelihood of continuing the word task (41.58%) relative to the control condition (26.21%), $\chi^2(1) = 5.38$, $p = 0.02$, 95% CI$_{difference}$ = [2.54%, 28.20%], Cohen's $h = 0.33$. However, among participants who were led to view their past performance as strong, a reset significantly decreased their likelihood of staying in the word task (39.00%) relative

to the control condition (60.58%), $\chi^2(1) = 9.49$, $p = 0.002$, 95% CI$_{difference}$ = [8.18%, 34.98%], Cohen's $h = 0.44$. Fig. 4 Panel A depicts the results.

5.2.2. Future performance

A two-way ANOVA on the number of words participants generated after the first 12 trials showed a significant interaction between past performance and reset manipulations, $F(1, 404) = 13.60$, $p < 0.001$, $\eta_p^2 = 0.03$, consistent with H1b. Specifically, when participants were led to view their past performance as weak, a reset increased output ($M = 12.77$ words, $SD = 16.92$) relative to the control condition ($M = 7.27$ words, $SD = 13.35$), $F(1, 404) = 6.11$, $p = 0.01$, 95% CI$_{difference}$ = [1.30, 9.70], $d = 0.36$. However, when participants were led to view their past performance as strong, a reset decreased output ($M = 11.33$ words, $SD = 16.84$) relative to the control condition ($M = 17.43$ words, $SD = 16.22$), $F(1, 404) = 7.52$, $p = 0.006$, 95% CI$_{difference}$ = [1.54, 10.67], $d = 0.37$. Fig. 4 Panel B depicts the results.

5.2.3. Self-efficacy

A two-way ANOVA on self-efficacy revealed a significant interaction between past performance and performance reset manipulations, $F(1, 404) = 19.61$, $p < 0.001$, $\eta_p^2 = 0.05$, consistent with H2. The planned contrast analysis showed that when participants were led to view their past performance as weak, a reset increased self-efficacy ($M = 4.14$, $SD = 2.36$) relative to the control condition ($M = 3.08$, $SD = 1.95$), $F(1, 404) = 12.72$, $p < 0.001$, 95% CI$_{difference}$ = [0.46, 1.66], $d = 0.49$; when participants were led to view their past performance as strong, a reset decreased self-efficacy ($M = 5.39$, $SD = 2.24$) relative to the control condition ($M = 6.19$, $SD = 1.93$), $F(1, 404) = 7.27$, $p = 0.007$,
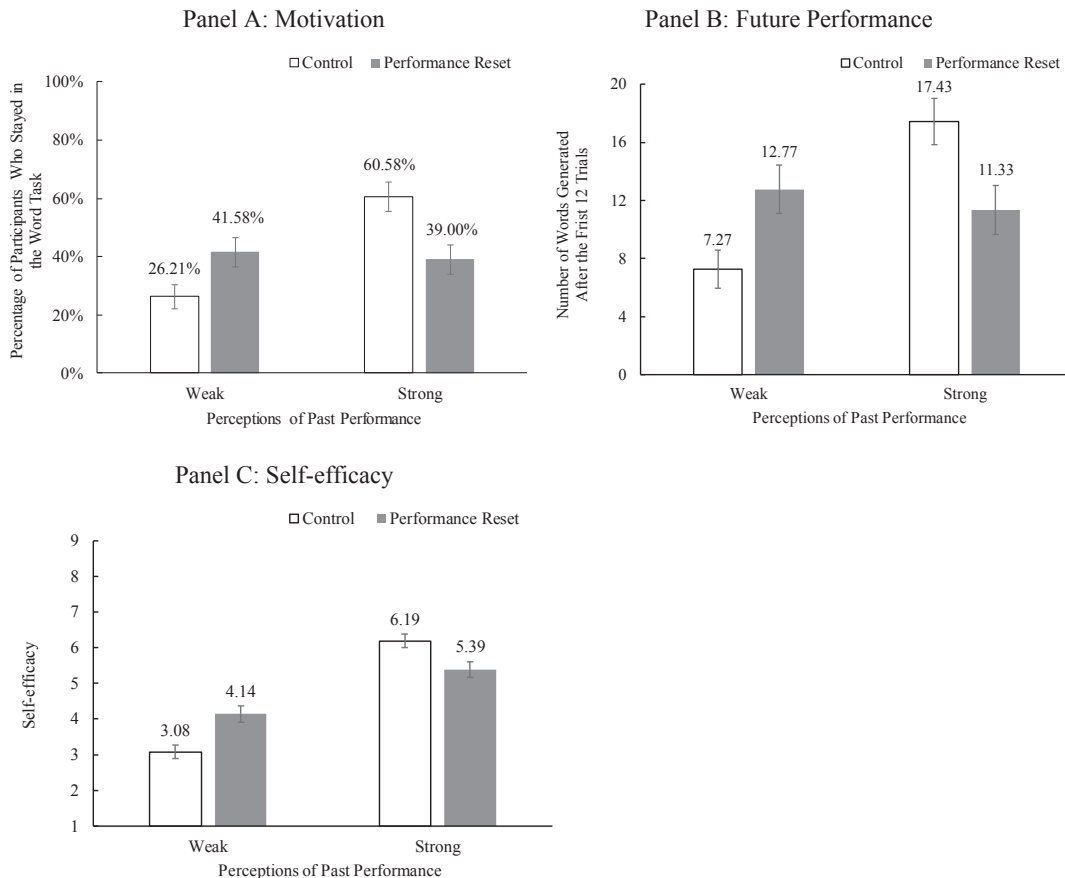
Note. Error bars represent the standard error of the mean.

Fig. 4. Motivation, future performance, and self-efficacy as a function of performance reset and past performance manipulations (Study 3). Panel A: Motivation. Panel B: Future Performance. Panel C: Self-efficacy. Note. Error bars represent the standard error of the mean.

95% CI$_{difference}$ = [0.23, 1.38], d = 0.38.

5.2.4. Mediation by self-efficacy

Next, I examined the role of self-efficacy in driving the relationship between a reset, past performance, and motivation (H3a) as well as future performance (H3b) using PROCESS model 8 (Hayes, 2013). Regarding motivation, a 5000-sample bootstrap analysis estimated an indirect effect of the interaction between reset and past performance manipulations via self-efficacy as −0.29 (SE = 0.12; 95% CI = [−0.56, −0.11]). When past performance was weak, the indirect effect of a reset on motivation via self-efficacy was positive and significant (indirect effect = 0.16; SE = 0.07; 95% CI = [0.05, 0.35]); when past performance was strong, the indirect effect of a reset on motivation via self-efficacy was negative and significant (indirect effect = −0.12; SE = 0.06; 95% CI = [−0.28, −0.03]). Thus, H3a was supported. Regarding future performance, a 5000-sample bootstrap analysis estimated an indirect effect of the interaction between reset and past performance manipulations via self-efficacy as −2.17 (SE = 0.93; 95% CI = [−4.28, −0.63]). When past performance was weak, the indirect effect of a reset on future performance via self-efficacy was positive and significant (indirect effect = 1.23; SE = 0.59; 95% CI = [0.30, 2.59]); when past performance was strong, the indirect effect of a reset on future performance via self-efficacy was negative and significant (indirect effect = −0.93; SE = 0.48; 95% CI = [−2.01, −0.19]). Thus, H3b was supported.

5.2.5. Alternative explanations

Two-way ANOVAs revealed that reset and past performance manipulations interacted to predict the perceived importance of building a performance streak ($F(1, 404)$ = 8.88, p = 0.003, η_p^2 = 0.02), positive affect ($F(1, 404)$ = 36.55, p < 0.001, η_p^2 = 0.08), and negative affect ($F(1, 404)$ = 15.87, p < 0.001, η_p^2 = 0.04). Notably, when I tested measures of alternative explanations along with self-efficacy as potential mediators (Hayes, 2013), self-efficacy was the only significant mediator of the interactive effect of reset and past performance on motivation (indirect effect = −0.26; SE = 0.15; 95% CI = [−0.62, −0.03]) and future performance (indirect effect = −2.65; SE = 1.19; 95% CI = [−5.36, −0.64]). Regarding motivation, the indirect effect of a reset via self-efficacy was 0.15 (SE = 0.09; 95% CI = [0.02, 0.38]) when past performance was weak and −0.11 (SE = 0.07; 95% CI = [−0.31, −0.01]) when past performance was strong. Regarding future performance, the indirect effect of a reset via self-efficacy was 1.51 (SE = 0.75; 95% CI = [0.30, 3.19]) when past performance was weak and −1.14 (SE = 0.61; 95% CI = [−2.56, −0.17]) when past performance was strong. See Appendix C for how measures of alternative explanations differed by condition and the indirect effects via them.

5.3. Discussion

Study 3's design mimicked common situations in which employees choose to either keep working on a project or switch to a different one (Dai, Dietvorst, Tuckfield, Milkman, & Schweitzer, 2017). Participants were rewarded based on their absolute performance and were incentivized to choose a task that could maximize their payoff. However, their motivation to exert effort in the focal task and their subsequent performance on the task were affected by the mere separation of past and current performance in the feedback interface, as well as by their perceptions of past performance (H1a and H1b). Also, Study 3 shows that self-efficacy mediated the effects of a reset and past performance on motivation and future performance (H2, H3a, and H3b), whereas alternative explanations (including the perceived importance of keeping a good streak, positive affect, and negative affect) did not explain my results.

6. Study 4: Field study of major league baseball players

In Study 4, I conducted an archival study about a field setting: Major League Baseball (MLB), a professional baseball organization in the United States and Canada. Sport settings have often been used to study organizational behavior, particularly motivation and performance (e.g., Hofmann, Jacobs, & Gerras, 1992; Flynn & Amanatullah, 2012; Marr & Thau, 2014; Swaab, Schaerer, Anicich, Ronay, & Galinsky, 2014) because of the availability of high-resolution, longitudinal performance data, the high-stakes nature of individual and team performance, and the sufficient contextual overlap between sport and other organizational settings (see Day, Gordon, & Fink, 2012 for a review).

6.1. Setting

MLB offers an appropriate setting for studying the effects of performance resets for several reasons. First, the MLB consists of two leagues: the American League (AL) and the National League (NL). When a player is traded across leagues during the regular season, his season-to-date statistics are reset: his statistics in the new league will be tracked from a clean slate, independent of his statistics in the previous league. However, if a player is traded to another team within the same league, his statistics will be continuously tracked from his statistics before the trade. As an example, imagine a player whose season-to-date batting average right before a trade is 0.275. If he is traded across leagues, his batting average at his first post-trade game will start from 0.000, and his pre-trade batting average will be kept at 0.275; if he is traded within the same league, his batting average at his first post-trade game will start at 0.275. In both cases, his statistics prior to the trade are included in the calculation of his overall seasonal statistics and his career statistics. Another advantage of the MLB context is that baseball is "an individual sport masquerading as a team sport" (Simmons, 2012), meaning that each team member performs more or less on his own (Swaab et al., 2014). Individual performance, particularly batting performance, is reasonably independent of teammates' efforts (Lord & Hohenfeld, 1979; Mandelbaum, 2005).

In MLB, whether a player knows in advance that he will be traded depends on several factors, including whether he is traded after MLB's non-waiver trade deadline (only 30% of regular-season trades in my data happened after this deadline) and whether he has the right to veto trades to another team (due to a no-trade clause in his contract or his long tenure in MLB and his current team). More often than not, players do not know in advance whether and when they will be traded. Thus, this study primarily examines the effects of unanticipated performance resets.

In this study, I focus on batting performance (see Appendix E for reasons). Compared to other batting statistics, I expect the reset on batting average—defined as a batter's number of hits divided by his number of batting attempts (at bats)—following a cross-league trade to be particularly salient to players. Batting average is deemed "easily the most common statistic in baseball and the most understood" (The Baseball Almanac, n.d.). For a long time, batting average was one of the key components by which players were judged (Lewis, 2003). Game box scores at minimum include batting averages, and the player at bat has his batting average displayed on the scoreboard and on television. Furthermore, players track their own batting averages closely and modify their behavior to reach a desirable batting-average threshold (Pope & Simonsohn, 2010).

Players with high batting averages early in the season may feel less motivated when their batting averages are reset due to a cross-league trade; in contrast, players with low batting averages early in the season may feel more motivated by the same reset. I test H1b by investigating whether performance resets due to cross-league trades differentially influence players' post-trade batting performance, depending on their batting averages prior to a trade. Importantly, to account for the effect of switching teams (Bateman, Karwan, & Kazee, 1983) and the potential

influence of regression to the mean, I constructed a "control" condition using players who were traded within the same league during the regular season and thus did not experience a reset. Notably, teams in a league are not geographically concentrated. A player has to travel a similar distance if he is traded across leagues (median = 1167 miles, mean = 1335 miles, SD = 855 miles) as he would if he were traded within the same league (median = 1199 miles, mean = 1396 miles, SD = 845 miles), $t(433) = 0.74$, $p = 0.46$.

6.2. Data

Play-by-play data for all MLB players from 1975 to 2014 regular seasons, as well as correlative information about trades, was obtained from Retrosheet's (n.d.) event files and transaction database. Since each player may appear in multiple seasons in this data, I use the term *player-season* to refer to a given player in a given season. I identified 1404 player-seasons in which a player batted in at least one game before and after a regular season trade, and had not switched teams in that season prior to the trade in question.

As explained above, the reset on batting average following a cross-league trade is particularly salient to players. Thus, to quantify a given batter's past performance, I first calculated his batting average using statistics from all games played for the team from which he was traded (hereafter, *pre-trade batting average*). To ensure sufficiently granular and meaningful pre-trade batting averages, I restricted my main analysis to players who had more than 100 at bats during the pre-trade period, consistent with other research.[4] Thus, my main analysis relied on 701 player-seasons involving 269,623 observations tracking each batter's at bats. Among those 701 trades, 42.23% were cross-league trades.

6.3. Variables

6.3.1. Dependent variable

The *hit indicator* was a dummy variable that equaled 1 if a player successfully got a hit at bat and 0 if the player did not get a hit at bat ($M = 0.262$, $SD = 0.440$).

6.3.2. Independent variables

- The *performance-reset indicator* was a dummy variable that equaled 1 if a player was traded across leagues and 0 if the player was traded within the same league ($M = 0.420$, $SD = 0.494$). In this study, cross-league trades were the "treatment" that induced performance resets, while within-league trades served as the "control" without performance resets.
- To operationalize past performance, I used league average—the average of batting averages across all players in a given season in the league from which a player was traded—as the benchmark to assess whether the player achieved strong or weak performance prior to the trade, given that batting averages vary by league from

season to season. Specifically, for each player-season, *past performance* equaled the difference between the player's pre-trade batting average and the corresponding league average, with higher values indicating stronger batting performance during the pre-trade period of the season ($M = -0.003$, $SD = 0.031$).

- For each observation of at bat, the *post-trade indicator* equaled 1 if the at bat observation occurred after the player in question was traded in a given season, and 0 if the at bat observation occurred before the player was traded in that season ($M = 0.368$, $SD = 0.482$).

6.3.3. Control variables

My analysis controlled for a number of factors that might influence batting performance. First, batters traded across leagues may face pitchers whom they have not encountered, while batters traded within the same league continue to bat against the same set of pitchers. Though the direction of the effects of unfamiliarity between batters and pitchers on batting average is debatable, I addressed potential concerns about batter-pitcher familiarity by controlling for the number of times a batter encountered a given pitcher up to a given at bat. Second, teams differ in how favorable their ballparks are to batters (vs. pitchers), as indicated by their Batting Park Factor (BPF). I controlled for the BPF of the home team corresponding to each observation in my analysis (see Appendix F for details about my BPF measure). Third, it is plausible that players were traded to teams with different qualities and potentials based on the players' past performance and the type of trade. Thus, for each observation of at bat for a given player, I controlled for team performance by including both the percentage of games the player's team had won up to that date (*season-to-date winning percentage*) and the team's batting average up to that date (*season-to-date team batting average*). Additionally, I controlled for the calendar month corresponding to each observation. My results are robust to excluding these control variables (Appendix G). Lastly, I included player-season fixed effects to control for any individual differences that did not change for each player in a season (Wooldridge, 2010). Table 1 displays the descriptive statistics of and correlations among my dependent, independent, and control variables.

6.4. Analysis strategy

I relied on the following OLS regression specification to predict the hit indicator for at bat i for player-season j:

$$
\begin{aligned}
hit\ indicator_{ij} = {} & \alpha + \beta_1 (performance-reset\ indicator)_j \\
& + \beta_2 (past\ performance)_j + \beta_3 (post-trade\ indicator)_{ij} \\
& + \beta_4 [(performance-reset\ indicator)_j \\
& \times (past\ performance)_j] \\
& + \beta_5 [(performance-reset\ indicator)_j \\
& \times (post-trade\ indicator)_{ij}] + \beta_6 [(past\ performance)_j \\
& \times (post-trade\ indicator)_{ij}] \\
& + \beta_7 [(performance-reset\ indicator)_j \\
& \times (past\ performance)_j \times (post-trade\ indicator)_{ij}] + X_{ij} \\
& + \delta_j + \varepsilon_{ij},
\end{aligned}
$$

where the performance-reset indicator, past performance, post-trade indicator, their two-way and three-way interactions, as well as control variables described in the previous section (denoted by X_{ij}) were included. δ_j represents player-season fixed effects,[5] and ε_{ij} is the error

---

[4] When a player only has a small number of at bats, he could have an extreme batting average that reflects noise. For example, in my data, pre-trade batting averages range from 0.000 to 1.000 if I do not impose any exclusion criterion, whereas pre-trade batting averages have a much more reasonable range among players with at least a certain number of at bats (e.g., 0.151–0.337 for 100 at bats and 0.103–0.337 for 50 at bats). Thus, restricting my analysis to players with at least a certain number of at bats reduces the likelihood that random variation produces outliers and drives my results. I report results based on 100 at bats because 100 at bats fall within the range of the thresholds used by past research (e.g., 50 in Bateman et al., 1983; 150 in Hofmann et al., 1992; 200 in Pope & Simonsohn, 2010), allowing me to balance the need of (a) having a meaningful measure of past performance and (b) having a sufficient number of players from the statistical power perspective. Importantly, my results are robust if I do not have any restrictions or adopt various exclusion criteria used by past research (see Appendix G).

[5] Any variables that had a fixed value for a given player-season (e.g., the performance-reset indicator and past performance) were collinear to player-season fixed effects and cannot be estimated along with player-season fixed

Table 1
Summary statistics of and correlations among study variables (Study 4).

Variable	Mean	SD	1	2	3	4	5	6	7	8
1. Hit indicator	0.262	0.440								
2. Performance-reset indicator	0.420	0.494	0.000							
3. Past performance	−0.003	0.031	0.051[*]	−0.042[*]						
4. Post-trade indicator	0.368	0.482	0.005[*]	−0.017[*]	−0.038[*]					
5. Batter-pitcher encounters	9.391	11.269	0.007[*]	−0.074[*]	0.068[*]	−0.063[*]				
6. Batting park factor	100.084	2.376	0.019[*]	−0.007[*]	0.039[*]	−0.010[*]	−0.011[*]			
7. Month	6.492	1.696	0.003	−0.010[*]	0.012[*]	0.766[*]	−0.040[*]	−0.006[*]		
8. Season-to-date team batting average[a]	0.260	0.019	0.012[*]	0.042[*]	0.101[*]	0.077[*]	−0.025[*]	0.105[*]	0.105[*]	
9. Season-to-date winning percentage[a]	0.490	0.110	−0.001	0.025[*]	−0.027[*]	0.174[*]	−0.027[*]	−0.024[*]	0.149[*]	0.430[*]

Note. N = 269,623 at bats.
[*] p-value < .05.
[a] Statistics that are relevant to season-to-date team batting average and season-to-date winning percentage are calculated based on 269,588 at bats (rather than the full sample of 269,623 at bats) because these variables have missing values for any team's first game in a season.
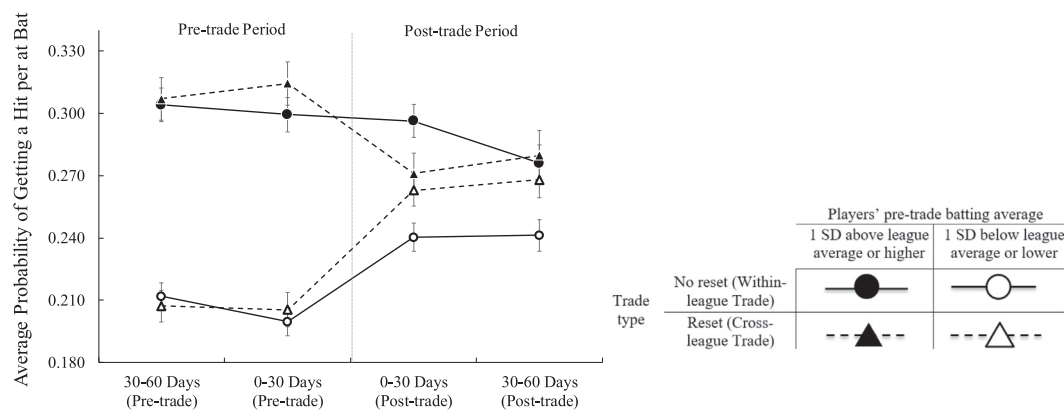
Note. Error bars correspond to standard errors. "1 SD" in the legend means one standard deviation.

Fig. 5. Major League Baseball players' average probability of getting a hit per at bat before and after regular season trades (Study 4). *Note.* Error bars correspond to standard errors. "1 SD" in the legend means one standard deviation.

term. I report results based on the ordinary least squares regression model in the paper because I control for a large number of fixed effects, and logistic regression models with fixed effects often produce inconsistent estimates unless the data meet stringent assumptions (i.e., the "incidental parameter problem"; Wooldridge, 2010). However, my results are robust if I use logistic regressions (Appendix G). To account for the non-independence of observations within each player-season, I clustered standard errors at the player-season level (Petersen, 2009). In another way to address the nesting nature of my data, I conducted a multilevel mixed-effect model with player-season random effects (Hofmann, 1997) and confirmed the robustness of my results (Appendix G).

6.5. Results

6.5.1. Analyses based on summary statistics

For ease of exposition, I first graphically present the results in Fig. 5. Approximately 90% of players in my sample had observations two months after they were traded before the regular season ended, but only around 50% had observations three months after their trade. Thus, Fig. 5 focuses on observations that happened within roughly 2 months

(footnote continued)
effects. Thus, the performance-reset indicator, past performance, and their interaction were not reported in Table 2. It is appropriate to control for player-season fixed effects because my regression analysis involves a difference-in-difference model and there are multiple observations within each player-season (Wooldridge, 2010). However, my results are robust if I (a) exclude player-season fixed effects or (b) use player-season random effects (Appendix G).

(60 days) before or after a trade, and it separates both the pre-trade and post-trade periods into two 30-day intervals. Fig. 5 depicts the average probability of getting a hit per at bat as a function of (a) whether the player exhibited strong or weak performance before the trade, (b) whether the trade was a cross-league trade (with a reset) or within-league trade (without a reset), and (c) each of the 30-day intervals prior to versus after the trade. Fig. 5 identifies players with strong past performance as those whose pre-trade batting averages were one standard deviation above league average or higher, and players with weak past performance as those whose pre-trade batting averages were one standard deviation below league average or lower. Note that for both players with strong past performance and players with weak past performance, the probability of getting a hit did not significantly differ between cross-league trades and within-league trades during the two 30-day intervals *before* trades (all *p*'s > 0.27).

My focus was on examining batting performance after players were traded. For players whose pre-trade batting averages were one standard deviation below league average or lower, their average probability of getting a hit per at bat was significantly higher following cross-league trades than following within-league trades during both 0–30 days ($M_{cross-league} = 0.263$, $SD = 0.440$ vs. $M_{within-league} = 0.240$, $SD = 0.427$, $t(7,249) = 2.22$, $p = 0.03$, 95% CI$_{difference}$ = [0.003, 0.043], $d = 0.05$) and 30–60 days ($M_{cross-league} = 0.268$, $SD = 0.443$ vs. $M_{within-league} = 0.241$, $SD = 0.428$, $t(5,779) = 2.33$, $p = 0.02$, 95% CI$_{difference}$ = [0.004, 0.049], $d = 0.06$) after a trade. However, the effects of performance resets induced by cross-league trades had the opposite effect among players whose pre-trade batting averages were one standard deviation above league average or higher. Specifically, their average probability of getting a hit per at bat during the period of

Table 2
Results of regressions predicting probability of getting a hit at an at bat (Study 4).

Dependent measure: Hit indicator	Trades that occurred during the regular season Model 1[b]	Trades that occurred between seasons Model 2[c] (Placebo test)
Post-trade indicator	0.010**	0.004**
	(0.003)	(0.002)
Performance-reset indicator × Post-trade indicator	−0.002	−0.002
	(0.004)	(0.002)
Past performance × Post-trade indicator	−0.538***	−0.672***
	(0.069)	(0.053)
Performance-reset indicator × Past performance × Post-trade indicator	−0.325**	0.024
	(0.101)	(0.077)
Control variables[a]	Yes	Yes
Observations	269,588	523,626
Number of player-seasons (fixed effects included)	701	753
R^2	0.0051	0.004

Note. Standard errors are clustered at the player-season level. *, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively.

[a] For each at bat, control variables include the number of times the batter faced the pitcher up to that at bat, batting park factor associated with the home team, performance of the team where the batter played (including the team's season-to-date batting average and season-to-date winning percentage), as well as calendar month.

[b] This model does not include observations from a team's first game in the season because two control variables, season-to-date team batting average and season-to-date winning percentage, have missing values for a team's first game in the season. Thus, the number of observations in this model is smaller than the full observations in my sample.

[c] Model 2 relies on 753 cases where a player (a) batted in the MLB both before and after he was traded between seasons and (b) had more than 100 at bats in the season right before he was traded while he was in the team from which he was traded.

0–30 days after a trade was significantly lower following cross-league trades ($M = 0.271$, $SD = 0.445$) than following within-league trades ($M = 0.296$, $SD = 0.457$), $t(5,307) = 1.98$, $p = 0.048$, 95% $CI_{difference}$ = [0.0002, 0.050], $d = 0.06$. This difference disappeared 30–60 days post-trade ($M_{cross-league} = 0.280$, $SD = 0.449$ vs. $M_{within-league} = 0.276$, $SD = 0.447$), $t(3,963) = 0.25$, $p = 0.80$, 95% $CI_{difference}$ = [−0.033, 0.025], $d = 0.01$. In summary, Fig. 5 provides evidence consistent with H1b, based on raw data. To control for potential alternative explanations (e.g., differences in team qualities) and take into account repeated observations within each player-season, I next turned to regression analyses.

6.5.2. Regression analyses

Model 1 in Table 2 presents the results from the OLS regression that predicted the hit indicator as described in the Analysis Strategy section. The positive coefficient on the post-trade indicator ($B = 0.010$, $SE = 0.003$, $p = 0.003$, 95% CI = [0.003, 0.017], $\eta_p^2 = 0.00003$) means that the probability of getting a hit on average increased by 1 percentage point after *within-league trades* without resets. The statistically insignificant two-way interaction between post-trade and performance-reset indicators ($B = -0.002$, $SE = 0.004$, $p = 0.53$, 95% CI = [−0.009, 0.005], $\eta_p^2 = 1.23e−06$) suggested that resets associated with cross-league trades did not significantly affect how batting averages changed post trade, when the measure of past performance equaled zero (i.e., when a player's pre-trade batting average equaled the league average).[6] Further, the negative interaction between past

[6] These results are consistent with Bateman et al. (1983) which found that batting averages increased after both cross-league and within-league trades during a regular season. Bateman et al. (1983) was interested in the basic effect

performance and the post-trade indicator ($B = -0.538$, $SE = 0.069$, $p < 0.001$, 95% CI = [−0.674, −0.402], $\eta_p^2 = 0.0002$) suggests that the lower a player's pre-trade performance, the more his hitting probability improved after a trade. This may simply reflect regression to mean and is irrelevant to my hypothesis.

The key variable of interest is the three-way interaction between the performance-reset indicator, past performance, and the post-trade indicator. The negative and significant three-way interaction ($B = -0.325$, $SE = 0.101$, $p = 0.001$, 95% CI = [−0.524, −0.127], $\eta_p^2 = 0.00003$) indicates that a player's past performance moderates the effects of a performance reset on post-trade performance. To illustrate the significant three-way interaction in Fig. 6, I plotted the regression-estimated change in the likelihood of getting a hit before versus after a trade, as a function of (a) whether the trade triggered a performance reset (i.e., a cross-league trade) or not (i.e., a within-league trade) and (b) whether the player had weak or strong past performance (i.e., pre-trade batting average one standard deviation below or above league average). When players' pre-trade batting averages were one standard deviation below their league average, cross-league trades and within-league trades were estimated to *increase* hit probability by 3.8 percentage points ($p < 0.001$, 95% CI = [0.029, 0.046]) and 2.9 percentage points ($p < 0.001$, 95% CI = [0.021, 0.037]), respectively, as compared to the pre-trade period. Thus, the performance resets induced by cross-league trades were estimated to significantly improve hit probability relative to the no-reset within-league trades by an additional 0.9 percentage points ($p = 0.044$, 95% CI = [0.0003, 0.0177])—equivalent to 3% of the average hit probability (i.e., 0.262) in my sample. However, when players' pre-trade batting averages were one standard deviation above their league average, cross-league trades and within-league trades were estimated to *decrease* hit probability by 2.2 percentage points ($p < 0.001$, 95% CI = [−0.031, −0.012]) and 0.8 percentage points ($p = 0.07$, 95% CI = [−0.017, 0.001]), respectively, as compared to the pre-trade period. Thus, the performance resets induced by cross-league trades were estimated to significantly decrease hit probability relative to the no-reset within-league trades by an additional 1.3 percentage points ($p = 0.014$, 95% CI = [−0.024, −0.003])—equivalent to 5% of the average hit probability (i.e., 0.262) in my sample. These findings, consistent with the patterns observed in Fig. 5 as well as H1b, suggest that cross-league trades associated with performance resets can improve hit probability more than within-league trades can when players' pre-trade batting performance was weak; however, cross-league trades associated with performance resets may lead to worse performance than within-league trades when players' pre-trade batting performance was strong.

### 6.5.3. Alternative explanations

Additional differences between cross-league and within-league trades (beyond the reset of players' statistics) may give rise to alternative explanations, which I addressed in several ways. First, I confirmed that cross-league trades and within-league trades in my sample were comparable on many relevant dimensions (e.g., performance of pre-trade and post-trade teams, how many at bats players had before and after they were traded; Appendix E). Second, as explained in the Control Variables section, my regression analyses controlled for a number of factors that may differ between the two types of trades (e.g., team performance, the familiarity between a batter and a pitcher).

Furthermore, I conducted one "placebo test" by leveraging the fact that a new season creates a reset on every player's performance regardless of any trades, meaning that cross-league trades and within-league trades occurring *between* seasons are equal in terms of resets. If

(footnote continued)
of job transitions, but did not examine the effects of performance resets that are uniquely introduced by cross-league trades or the moderating effects of players' past performance prior to a trade.
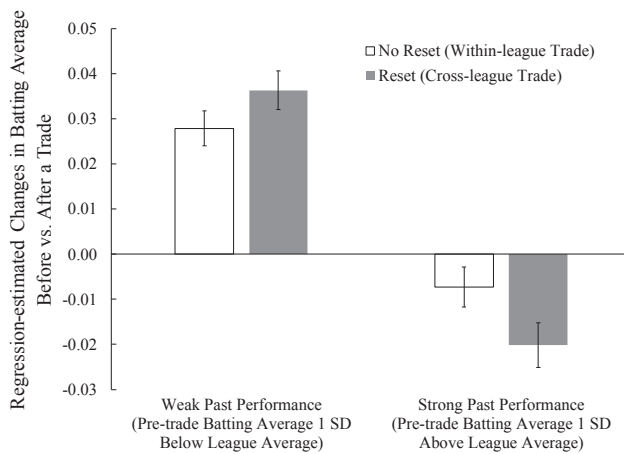
Fig. 6. Regression-estimated change in the probability of getting a hit per at bat before and after regular season trades as a function of performance reset and past performance (Study 4). *Note.* Error bars correspond to standard errors. "1 SD" means one standard deviation.

performance resets are largely responsible for my results comparing cross-league and within-league trades during the regular season, then trades occurring between seasons should *not* show the same results. I used the same regression model described in the Analysis Strategy section to analyze trades that occurred between seasons (see Appendix F for sample construction). Indeed, I did not see the same patterns predicted by H1b: There was not a significant three-way interaction between trade type, past performance, and the post-trade indicator (Model 2 in Table 2). This suggests that the interactive effect of regular season cross-league trades (relative to within-league trades) and players' past performance on post-trade performance is unlikely to be simply explained by other differences between the two types of trades besides the presence of performance resets (e.g., the difference in the designated-hitter rule).[7]

6.5.4. Robustness checks

In addition to the robustness checks that have been described earlier, my results in Model 1 of Table 2 were also robust if I only included players who had one trade during a regular season or if I used the absolute value of pre-trade batting average (instead of the difference between pre-trade batting average and league average) to operationalize players' past performance. Additionally, my results were robust when I added players who were *not* traded during a regular season and used these players' data to more precisely estimate the effects of control variables. See Appendix G for results about all robustness checks.

---

[7] The AL and NL use the same set of rules, except that the AL allows teams to have one player (known as the designated hitter) bat in place of the pitcher but the NL does not allow so. It is unclear how the difference in this designated-hitter (DH) rule between the AL and NL explains my findings that players with strong vs. weak pre-trade performance react differently to a cross-league trade (vs. a within-league trade). Nevertheless, I explored this as a potential alternative explanation. First, one implication of the DH rule is that it allows good *pitchers* who are bad at hitting to remain in the game in the AL. My results are robust if I exclude players who pitched in a season (Appendix G). Second, I confirmed that whether the initial league a player was traded from was the AL or NL did not significantly moderate my observed relationship between trade type, past performance, and post-trade performance (Appendix F). Also, my results were similar when I separately examined trades that originated from the AL vs. NL (Appendix F). Furthermore, if the DH rule is the primary cause of my observed effects, I should see the same effects in my "placebo test", but it was not the case as I described above.

6.6. Discussion

I found that the effects of performance resets induced by cross-league trades (vs. within-league trades without resets) on post-trade batting averages depend on players' performance prior to the trade. The same patterns did not emerge for trades that occurred between seasons when cross-league and within-league trades were equal in terms of performance resets, providing further evidence for the potential role played by performance resets during the regular season. These results cannot be explained by the potential effects of temporal landmarks because being traded to a new team is a temporal landmark regardless of where and when players are traded.

Although I addressed alternative explanations in multiple ways, whether a player was traded across leagues or within the same league was not randomly assigned. As is the case with all archival studies, I am cautious about drawing causal inferences from Study 4 alone; however, the robustness of my results across this archival study and the three experiments increases my confidence in the causal effects of performance resets. Further, Study 4 demonstrates the generalizability of my findings to a real-world context involving high stakes.

In this setting, similar to Studies 1–3, a player's pre-trade performance is *not* discarded once a reset happens. Instead, a player's pre-trade performance is visible to the public and remains relevant to his overall statistics season- and career-wise. However, the "treatment" of performance resets here is not purely based on framing (i.e., only affecting the tracking of performance) as in Studies 1–3, and may have practical implications for players. One practical consequence is that resets may affect players' chances of earning the "batting champion" title. Each league determines its batting champion at the end of each season, rewarding the player with the highest batting average and at least 502 plate appearances in the particular league that season. Though a player who is traded across leagues could still earn the title in his pre-trade league based on his pre-trade batting average, it would be challenging for him to have reached 502 plate appearances in the previous league before being traded, as trades tend to happen in the middle of the season. For a player with a very high batting average prior to a trade, a reset may demotivate him because he does not have enough plate appearances to earn the title in his previous league and is also unlikely to have enough plate appearances to earn the title in his new league even if he achieves a high batting average there. That said, this economic consideration should only matter to a small number of players with very high pre-trade batting averages. From 1974 to 2014, batting champions in the AL and NL had an average of 0.349 batting average and a minimal of 0.313 batting average. If I exclude players whose pre-trade batting average suggests that they had a chance of earning the batting title (e.g., 27 players whose pre-trade batting average above 0.310 or 16 players above 0.320), my results are robust (Appendix F). Also, this practical consideration does not explain why a reset increases performance among players with low pre-trade batting averages. Taken together, resets triggered by cross-league trades may be, to some extent, psychological, though I am cautious about arguing that economic considerations do not contribute to my findings.

7. General discussion

Across three laboratory experiments and one field study, I demonstrated that the effects of unanticipated performance resets on motivation and future performance varied as a function of past performance, with self-efficacy as one underlying mechanism. Specifically, unanticipated performance resets increased self-efficacy and generated motivational and performance benefits following weak past performance, but decreased self-efficacy and exacted motivational and performance costs following strong past performance. These patterns held when I objectively measured past performance (Studies 1 and 4) and when I experimentally manipulated past performance (Studies 2 and 3). By combining laboratory data and archival data, I both provide causal

evidence for my hypotheses and establish the external validity of my findings.

The size of the observed simple and interaction effects of resets and past performance on self-efficacy, motivation, and future performance was overall medium across my lab experiments, and is comparable to the effect size reported by recent papers that examine the interaction effects of two factors on motivation in the lab (e.g., Huang et al., 2017; Huang, Etkin, & Jin, 2017). In my field study, the proportion of variance in batting performance attributed to performance resets is very small, which is worth a cautious note. Nevertheless, when pre-trade batting averages are one standard deviation (SD) below (or above) league average, the estimated increase (or decrease) in batting averages associated with resets is 9 (or 13) points, which amounts to 3% (or 5%) of the average batting averages in my data and is non-trivial for professionals who are highly incentivized to perform well and are closely observed by the public.[8]

7.1. Theoretical and practical implications

My research makes several important theoretical contributions. First, as performance tracking technology becomes more prevalent, reports more granular data, and allows managers to increase the frequency of delivering and refreshing performance feedback, there is a growing need for researchers to understand the implications of different ways of tracking performance for individuals' motivation (Cappelli & Tavis, 2016). My research focuses on one aspect of performance tracking—whether future performance is decoupled from past performance. Performance resets often occur inside and outside of the workplace but have been overlooked as a potential antecedent of motivation. In the survey with full-time employees described in the introduction, 65% of respondents reported that their past performance still had a direct (26%) or indirect (39%) influence on their future job prospects (e.g., raises, promotions), and 35% reported that their past performance had no influence at all on their future job prospects after the reset in their example took place. To more cleanly capture the effect on motivation of the essence of performance resets (i.e., separate tracking of past and future performance), I have studied situations in which performance resets cause future performance to be tracked independently of past performance but do not discard past performance. Going beyond the psychological nature of performance resets, it would be valuable for future research to investigate how a reset affects motivation when it is accompanied by changes in the financial implications of past performance.

Second, a growing body of work has shown that people are more motivated to engage in positive behavior changes at transition points that feel like a fresh start (Dai et al., 2014, 2015; Hennecke & Converse, 2017) and events that prompt a fresh start mindset (Ayal & Gino, 2011; Price, Coulter, Strizhakova, & Schultz, 2018). My findings advance the field's understanding of how fresh starts influence behavior by studying a prevalent yet previously unexamined form of fresh start (i.e., performance resets), demonstrating a potential downside of fresh starts, and examining the effects of varying levels of past performance. The observed effects of performance resets hold whether temporal landmarks occur in all conditions (Studies 2 and 4) or did not occur in any

---

[8] In baseball, a 0.001 is considered a "point." As an example of non-trivial reset effects, in the last regular season in my data (2014), an increase in batting average by 9 points from 0.215 (1 SD below the AL's league average) to 0.224 means a jump in ranking by 26 among 245 players who had more than 100 at bats in the AL that season, and a decrease in batting average by 13 points from 0.287 (1 SD above the AL's league average) to 0.274 means a drop in ranking by 31. In the NL, an increase in batting average by 9 points from 0.213 (1 SD below the NL's league average) to 0.222 means a jump in ranking by 7 among 216 players with more than 100 at bats in the NL that season, and a decrease by 13 points from 0.285 (1 SD above the NL's league average) to 0.272 means a drop in ranking by 22.

condition (Studies 1 and 3). Also, I showed that self-efficacy drives both the positive and negative effects of performance resets on motivation and future performance, independent of alternative explanations (Studies 2 and 3).

Third, though my findings suggest that past performance moderates the effects of performance resets, another interpretation is that resets weaken the linkage between past performance and self-efficacy, which further influences motivation and subsequent performance. For example, in Study 3, without a reset, people who received positive performance feedback reported much higher self-efficacy ($M = 6.19$, $SD = 1.93$) than those who received negative performance feedback on the same objective performance ($M = 3.08$, $SD = 1.95$), $t(205) = 11.56$, $p < .0001$, 95% $CI_{difference} = [2.58, 3.65]$, $d = 1.61$. The difference in self-efficacy shrunk in the reset condition ($M_{strong\_past\_performance} = 5.39$, $SD = 2.24$ vs. $M_{weak\_past\_performance} = 4.14$, $SD = 2.36$), $t(199) = 3.85$, $p = .0002$, 95% $CI_{difference} = [1.89, 3.85]$, $d = 0.54$. Study 2 showed similar results. My findings identify a novel contextual factor—whether future performance is carried over or decoupled from past performance—that alters the strength of the well-established relationship between past performance and self-efficacy. In doing so, I add to the literature on how people adjust self-efficacy and motivation in response to feedback (e.g., Silver et al., 1995; Donovan & Williams, 2003; Tolli & Schmidt, 2008).

The results documented in this paper offer practical implications for managers. My findings suggest that given a chance to put past performance failures behind them, employees may avoid decreased self-efficacy and recover more easily from poor performance. Managers may help employees psychologically cope with negative feedback by offering them the option of resetting their performance statistics merely for the tracking purpose (e.g., allowing employees with low daily sales to track their performance in a new week from a clean slate). Employees given such an option may feel more confident and empowered. As an initial exploration of individuals' preferences for performance resets, I conducted a laboratory experiment in which participants had the option to reset their performance statistics in the same way as the reset treatment in Study 1 (see Appendix H). Participants were paid according to their total performance as in Study 1 and thus were incentivized to choose whether to reset their statistics based on what they believed would motivate them to a greater extent. I found that individuals were significantly more likely to choose the reset option if they were led to view their past performance as weak rather than strong. This suggests that people appreciate the opportunity to reset their statistics following weak past performance, even if the reset only affects how their performance is presented.

However, managers should be aware that performance resets affect employees differently depending on their past performance. In addition to simply reducing effort at work, the decrease in motivation among employees with strong past performance may also take subtler forms such as reallocating effort from the focal task to another task (Study 3), which may be costly to both individuals and organizations. Managers may consider communicating positive expectations to and instilling confidence in top performers when a reset occurs. By recognizing that performance resets do not affect all individuals equally, organizations and managers can better harness the benefits and avoid the disadvantages of resets.

Similar ideas may also apply outside of the workplace. For example, users of habit-tracking apps may fall prey to the so-called "what-the-hell" effect due to a recent lapse (Cochran & Tesser, 1996; Soman & Cheema, 2004) and become disengaged from the app (Silverman & Barasch, 2016). Instead of keeping the feedback interface constant over time, app designers may be able to keep users engaged by dynamically adjusting whether progress statistics are displayed in continuation of or independently from past performance, depending on users' recent progress. Also, while some habit-tracking apps (e.g., Productive and Strides) allow people to reset their statistics, others (e.g., Streaks) do not allow such a form of reset unless users close the account. For users

with recent failures, the reset option may boost their self-efficacy and engagement.

7.2. Limitations and directions for future research

One practical question is how long the patterns induced by resets persist. My laboratory experiments only measured motivation and performance over a short period, while my field study showed that a reset's effects on performance held for up to two months (Appendix F). Since performance resets may decrease high performers' motivation and even lead to their quitting, my findings have meaningful implications even if effects only persist in the short term. Future research may seek to examine the long-term effects of performance resets and identify the optimal frequency for resets to maximize their motivating impact on low performers.

While my research distinguishes primarily between strong and weak past performance and might seemingly suggest that a performance reset will not affect employees whose performance falls close to the average of a performance distribution, such a linear extrapolation may not be entirely accurate. An individual's perception of her past performance is influenced by other factors beyond the distribution of absolute performance, such as personal aspirations or a manager's expectations (Dai et al., 2017). Thus, even if an individual's past performance falls close to the average based on absolute measures, it is not necessarily true that the individual will be unaffected by a performance reset, based on her own perceptions of past performance. For example, in professional baseball, players may use other benchmarks in addition to league average to evaluate their performance in a given season, such as their statistics in the previous season. To explore this possibility, I replaced my measure of pre-trade performance (described in Study 4) with a new measure that equaled the difference between a player's pre-trade batting average and his batting average in the previous season. The new measure of pre-trade performance moderated the effects of performance resets on players' post-trade performance (Appendix F). Future research examining how different benchmarks influence employees' perceptions of their performance and change their responses to resets would be valuable.

In Studies 2 and 3, I presented participants with either steadily strong performance or steadily weak performance depending on their experimental condition in order to send unambiguous signals about participants' past performance and ensure the effectiveness of my past-performance manipulation. In Studies 1 and 4, patterns of past performance varied across individuals and were unsteady in most cases. For example, among professional baseball players in my main analysis, 85% players did not have steadily improving or declining batting averages in the months preceding their trade. Thus, complementing Studies 2 and 3, Studies 1 and 4 suggest that the reset effects are not limited to situations in which past performance was steadily good or bad. Yet, given that performance patterns are often noisy in the real world (as in Study 4) and the fluctuations in past performance may affect individuals' perceptions of their past progress, it would be valuable for future research to systematically explore how individuals with different patterns of past performance respond to resets.

Across my studies, the vividness and accessibility of past performance following a reset varied. For example, participants in Study 1 saw their average score in the first five trials while working on the last five trials, participants in Study 2 were not reminded of their past progress, and participants in Study 3 were not readily presented with their past performance but could click a link to view their past performance. My survey with full-time employees shows that the accessibility of past performance following a performance reset varies in the real world as well: 18% of respondents indicated that they could not access their past performance records anymore, 31% did not readily see their past records when receiving a new performance record but could find their past records somewhere if they intended to look them up, and 48% could easily see their past performance records when receiving a

new performance record. It would be useful to understand whether our observed reset effects are accentuated in cases where people are not reminded of their past records because the separation and contrast between past and future performance can be amplified in these cases.

Additionally, it may be worth examining how performance resets influence the views and reactions of others surrounding the focal actor. The results observed in this paper are unlikely to be driven by performance resets wiping off people's reputation associated with past performance: my lab studies conducted in private settings did not involve reputational concerns at all, and in my field study, baseball players' pre-trade performance is known to the public and their reputations carry over even after a cross-league trade (with a reset). If performance resets do affect reputation—making past performance less relevant to the present reputation of the individual—the effects of performance resets observed in this paper might become stronger.

Finally, my paper has focused on unanticipated performance resets. I expect my findings to extend beyond unanticipated, one-shot resets. For instance, as mentioned in the introduction, salespeople who were top sellers in the previous month feel demotivated when they have to start over in the new month, although they are aware of the routine reset in advance. The ways in which anticipated and recurrent resets affect motivation are open for further investigation. To help managers decide whether and when to reveal interventions that introduce a performance reset, future studies could explore how employees adjust their efforts both *before* and *after* an anticipated reset occurs. This will also contribute to research on the anticipatory effect of a fresh start (Alter & Hershfield, 2014; Hennecke & Converse, 2017; Koo, Mai, Dai, & Song, 2018).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.obhdp.2018.06.002.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Alter, A. L., & Hershfield, H. E. (2014). People search for meaning when they approach a new decade in chronological age. *Proceedings of the National Academy of Sciences, USA, 111,* 17066–17070. https://doi.org/10.1073/pnas.1415086111.

Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. In M. Mikulincer, & P. R. Shaver (Eds.). *The social psychology of morality: Exploring the causes of good and evil* (pp. 149–166). Washington, DC: American Psychological Association.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248–287. https://doi.org/10.1016/0749-5978(91)90022-L.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.). *Self-efficacy beliefs of adolescents* (pp. 120–145). Info Age Publishing: Greenwich, CT.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87–99. https://doi.org/10.1037/0021-9010.88.1.87.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*(3), 586–598. https://doi.org/10.1037/0022-3514.41.3.586.

Baseball Almanac (n.d.). Offensive stats 101. Retrieved March 17, 2015, from http://www.baseball-almanac.com/stats.shtml.

Bateman, T. S., Karwan, K. R., & Kazee, T. A. (1983). Getting a fresh start: A natural quasi-experimental test of the performance effects of moving to a new job. *Journal of Applied Psychology, 68*(3), 517–524. https://doi.org/10.1037/0021-9010.68.3.517.

Benartzi, S., & Lehrer, J. (2015). *The smarter screen: Surprising ways to influence and improve online behavior.* New York: Portfolio/Penguin.

Broemer, P., Grabowski, A., Gebauer, J. E., Ermel, O., & Diehl, M. (2008). How temporal distance from past selves influences self-perception. *European Journal of Social Psychology, 38*(4), 697–714. https://doi.org/10.1002/ejsp.469.

Brown, S. P., Jones, E., & Leigh, T. W. (2005). The attenuating effect of role overload on relationships linking self-efficacy and goal level to work performance. *Journal of Applied Psychology, 90*(5), 972–979. https://doi.org/10.1037/0021-9010.90.5.972.

Cappelli, P., & Tavis, A. (2016). The performance management revolution. *Harvard Business Review, 94*(10), 58–67.

Cochran, W., & Tesser, A. (1996). The "what the hell" effect: Some effects of goal proximity and goal framing on performance. In L. L. Martin, & A. Tesser (Eds.). *Striving and feeling: Interactions among goals, affect, and self- regulation.* Hillsdale, NJ: Psychology Press.

Dai, H., Dietvorst, B., Tuckfield, B., Milkman, K. L., & Schweitzer, M. E. (2017). Quitting when the going is tough: The downside of high performance expectations. *Academy of Management Journal.*. https://doi.org/10.5465/amj.2014.1045.

Dai, H., Milkman, K. L., & Riis, J. (2014). The fresh start effect: Temporal landmarks motivate aspirational behavior. *Management Science, 60*(10), 2563–2582. https://doi.org/10.1287/mnsc.2014.1901.

Dai, H., Milkman, K. L., & Riis, J. (2015). Put your imperfections behind you: Temporal landmarks spur goal pursuit when they signal new beginnings. *Psychological Science, 26*(12), 1927–1936. https://doi.org/10.1177/0956797615605818.

Day, D. V., Gordon, S., & Fink, C. (2012). The sporting life: Exploring organizations through the lens of sport. *Academy of Management Annals, 6*(1), 397–433.

Donovan, J. J., & Williams, K. J. (2003). Missing the mark: Effects of time and causal attributions on goal revision in response to goal-performance discrepancies. *Journal of Applied Psychology, 88*(3), 379–390. https://doi.org/10.1037/0021-9010.88.3.379.

Flynn, F. J., & Amanatullah, E. T. (2012). Psyched up or psyched out? The influence of coactor status on individual performance. *Organization Science, 23*(2), 402–415.

Hanko, K., Crusius, J., & Mussweiler, T. (2010). When I and me are different: Assimilation and contrast in temporal self-comparisons. *European Journal of Social Psychology, 40*(1), 160–168. https://doi.org/10.1002/ejsp.625.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Press.

Hennecke, M., & Converse, B. A. (2017). Next week, next month, next year: The role of temporal boundaries in resolving the dilemma of initiation. *Social Psychological and Personality Science, 8*(8), 918–926. https://doi.org/10.1177/1948550617691099.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*(6), 723–744. https://doi.org/10.1016/S0149-2063(97)90026-X.

Hofmann, D. A., Jacobs, R., & Gerras, S. J. (1992). Mapping individual performance over time. *Journal of Applied Psychology, 77*(2), 185–195.

Huang, S., Etkin, J., & Jin, L. (2017). How winning changes motivation in multiphase competitions. *Journal of Personality and Social Psychology, 112*(6), 813–837. https://doi.org/10.1037/pspa0000082.

Huang, S., Jin, L., & Zhang, Y. (2017). Step by step: Sub-goals as a source of motivation. *Organizational Behavior and Human Decision Processes, 141*, 1–15. https://doi.org/10.1016/j.obhdp.2017.05.001.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*(3), 352–376.

Imas, A. (2016). The realization effect: Risk-taking after realized versus paper losses. *The American Economic Review, 106*(8), 2086–2109. https://doi.org/10.1257/aer.20140386.

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research, 43*(1), 39–58. https://doi.org/10.1509/jmkr.43.1.39.

Koo M., Mai K.M., Dai H., Song C.E. (2018). An ironic effect of "fresh starts": How anticipated temporal landmarks undermine motivation for continued goal pursuit, Working Paper.

Lewis, M. (2003). *Moneyball.* New York: W. W. Norton & Company.

Li, X., & Soman, D. (2010). The effect of intermediate rewards on the effectiveness of incentive programs. Working paper. Singapore: The National University of Singapore.

Li, X., Wei, L., & Soman, D. (2010). Sealing the emotions genie: The effects of physical enclosure on psychological closure. *Psychological Science, 21*(8), 1047–1050. https://doi.org/10.1177/0956797610376653.

Libby, L. K., & Eibach, R. P. (2002). Looking back in time: Self-concept change affects visual perspective in autobiographical memory. *Journal of Personality and Social Psychology, 82*(2), 167–179. https://doi.org/10.1037//0022-3514.82.2.167.

Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task-performance. *Journal of Applied Psychology, 69*(2), 241–251. https://doi.org/10.1037/0021-9010.69.2.241.

Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*(1), 91–108. https://doi.org/10.1037/0033-295X.100.1.91.

Lord, R. G., & Hohenfeld, J. A. (1979). Longitudinal field assessment of equity effects on the performance of major league baseball players. *Journal of Applied Psychology, 64*(1), 19–26. https://doi.org/10.1037/0021-9010.64.1.19.

Maddox, K. B., Rapp, D. N., Brion, S., & Taylor, H. A. (2008). Social influences on spatial memory. *Memory & Cognition, 36*(3), 479–494.

Mandelbaum, M. (2005). *The meaning of sports.* New York: PublicAffairs.

Marr, J. C., & Thau, S. (2014). Falling from great (and not-so-great) heights: How initial status position influences performance after status loss. *Academy of Management Journal, 57*(1), 223–248.

Mishra, A., & Mishra, H. (2010). Border bias: The belief that state borders can protect against disasters. *Psychological Science, 21*(11), 1582–1586. https://doi.org/10.1177/0956797610385950.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 28*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30.

Nunes, J., & Drèze, X. (2006). The endowed progress effect: How artificial advancement increases effort. *Journal of Consumer Research, 32*(4), 504–512. https://doi.org/10.1086/500480.

O'Brien, J. (2015). Beginning of the month sales motivation: Dealing with the monthly sales reset (and not burning out). Accelerate blog, Hirevue. Retrieved from https://blog.hirevue.com/coach/first-of-the-month-blues-dealing-with-the-monthly-sales-reset-and-not-burning-out.

Peetz, J., & Wilson, A. E. (2013). The post-birthday world: Consequences of temporal landmarks for temporal self-appraisal and motivation. *Journal of Personality and Social Psychology, 104*(2), 249–266. https://doi.org/10.1037/A0030477.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22*(1), 435–480. https://doi.org/10.1093/rfs/hhn053.

Pope, D., & Simonsohn, U. (2010). Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychological Science, 22*(1), 71–79. https://doi.org/10.1177/0956797610391098.

Price, L. L., Coulter, R. A., Strizhakova, Y., & Schultz, A. E. (2018). The fresh start mindset: Transforming consumers' lives. *Journal of Consumer Research, 45*(1), 21–48. https://doi.org/10.1093/jcr/ucx115.

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty, 19*(1–3), 171–197. https://doi.org/10.1023/A:1007879411489.

Retrosheet (n.d.). Play-by-play data files (event files). Retrieved December 20, 2014 http://www.retrosheet.org/game.htm.

Ross, M., & Wilson, A. E. (2002). It feels like yesterday: Self-esteem, valence of personal past experiences, and judgments of subjective distance. *Journal of Personality and Social Psychology, 82*(5), 792–803. https://doi.org/10.1037/0022-3514.82.5.792.

Schmidt, A. M., & DeShon, R. P. (2010). The moderating effects of performance ambiguity on the relationship between self-efficacy and performance. *Journal of Applied Psychology, 95*(3), 572–581. https://doi.org/10.1037/a0018289.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*(3), 313–322. https://doi.org/10.1037/0022-0663.77.3.313.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*(1), 54–61. https://doi.org/10.1037/0022-0663.79.1.54.

Schwarz, N., & Bless, H. (2007). Mental construal processes: The inclusion/exclusion model. In D. A. Stapel, & J. Suls (Eds.). *Assimilation and contrast in social psychology* (pp. 119–141). New York: Psychology Press.

Seo, M. G., & Ilies, R. (2009). The role of self-efficacy, goal, and affect in dynamic motivational self-regulation. *Organizational Behavior and Human Decision Processes, 109*(2), 120–133.

Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin, 124*(3), 423–442. https://doi.org/10.1037/0033-2909.124.3.423.

Silver, W. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes, 62*(3), 286–299. https://doi.org/10.1006/obhd.1995.1051.

Silverman, J., & Barasch, A. (2016). You are what you track: The effect of failing to log an experience on future use of tracking apps. *Advances in Consumer Research, 44*, 21–25.

Simmons, B. (2012). A-Rod is a clubhouse guy? In a manner of speaking, yes? ESPN.com. Retrieved from http://sports.espn.go.com/espn/magazine/archives/news/ story?page=magazine-20090420-article3.

Soman, D., & Cheema, A. (2004). When goals are counterproductive: The effects of violation of a behavioral goal on subsequent performance. *Journal of Consumer Research, 31*, 52–62. https://doi.org/10.1086/383423.

Soster, R. L., Monga, A., & Bearden, W. O. (2010). Tracking costs of time and money: How accounting period affect mental accounting. *Journal of Consumer Research, 37*(4), 712–721. https://doi.org/10.1086/656388.

Stajkovic, A., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240.

Sullivan, B. A., O'Connor, K. M., & Burris, E. R. (2006). Negotiator confidence: The impact of self-efficacy on tactics and outcomes. *Journal of Experimental Social Psychology, 42*(5), 567–581.

Swaab, R. I., Schaerer, M., Anicich, E. M., Ronay, R., & Galinsky, A. D. (2014). The too-much-talent effect team interdependence determines when more talent is too much or not enough. *Psychological Science, 25*(8), 1581–1591. https://doi.org/10.1177/0956797614537280.

Tolli, A. P., & Schmidt, A. M. (2008). The role of feedback, causal attributions, and self-efficacy in goal revision. *Journal of Applied Psychology, 93*(3), 692–701. https://doi.org/10.1037/0021-9010.93.3.692.

Tu, Y., & Soman, D. (2014). The categorization of time and its impact on task initiation. *Journal of Consumer Research, 41*(3), 810–822.

Tversky, B. (1992). Distortions in cognitive map. *Geoforum, 23*(2), 131–138.

Watson, D., Clark, L. A., & Tellegan, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063.

Wilson, H. J. (2013). Wearables in the workplace. *Harvard Business Review, 91*(11), 27.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* Cambridge, MA: The MIT Press.