Title

Miscalibration, Missing Attributions, and Mispredictions: An Exploration of Momentum, Efficacy, and Performance Expectations

Author

Habbert, Rachel

Publication Date

2019

Peer reviewed|Thesis/dissertation

Miscalibration, Missing Attributions, and Mispredictions:
An Exploration of Momentum, Efficacy, and Performance Expectations

By

Rachel Emma Habbert

A dissertation to be submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Juliana Schroeder, Chair
Professor Don A. Moore
Professor Cameron Anderson
Professor Dacher Keltner

Spring 2019

Abstract

Miscalibration, Missing Attributions, and Mispredictions:
An Exploration of Momentum, Efficacy, and Performance Expectations

by

Rachel Emma Habbert

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Juliana Schroeder, Chair

Momentum and efficacy are important forces to understand for performance-related outcomes in a variety of contexts, from athletic contests to mental challenges to day-to-day tasks. This dissertation explores perceptions and consequences of those forces, in three chapters.

In Chapter 1, which includes seven total experiments (*N* = 3,052), I find that the experience of gaining momentum leads to increased performance expectations (and losing momentum to decreased expectations), but that these expectations are often miscalibrated as compared to actual performance. I further find that participants will bet on their performance expectations when they experience gaining momentum, even though experiencing momentum does not consistently lead to actual performance increases commensurate with expectations.

In Chapter 2, I discuss the results of four experiments (*N* = 1,347). I replicate the findings from Chapter 1 that perceived momentum leads to performance expectations, but I find that link only exists if the performer is judged to have control over the outcome of the contest (i.e., perceiving that the performer has efficacy). When efficacy is either attenuated or not present (e.g., in games of chance), observers no longer believe that gaining momentum will lead to improved performance.

In Chapter 3, I present nine experiments (*N* = 2,729) which uncover and then correct a misprediction about the effect of task ordering on performers' efficacy. I find that people predict that completing tasks in increasing-difficulty (vs. decreasing-difficulty) order will lead to greater felt efficacy and that people consequently prefer to complete tasks in increasing-difficulty order. I then expose that misprediction and find instead that there is either no effect of task ordering on efficacy, or that the opposite is true: completing tasks in decreasing-difficulty order leads to more reported efficacy. Finally, I explore a potential mechanism to correct the misprediction: clearer simulation of the actual experience of completing tasks in different orders.

Across these three chapters and 7,128 total participants, I demonstrate a strong link between momentum perceptions and performance expectations but find that these expectations are often miscalibrated as compared to reality. I also show that perceived momentum must signal efficacy in order to affect performance expectations. Finally, I present evidence that efficacy can be

generated by completing tasks in decreasing-difficulty order, despite people's expectations that completing tasks in increasing-difficulty order will increase perceived efficacy and their preferences for starting with their easiest task.

Table of Contents

Acknowledgements

This success belongs not just to me, but to all the people in my life who have helped me to become the person and the scholar that I am today. I could not have even begun to pursue this PhD without my family's unwavering support and belief in my abilities. I want to specifically thank my parents, who ensured that I received an excellent education and fostered an environment of learning and curiosity at home as well, and who continue to provide advice, vocal pride, and unconditional love. I also want to thank my sisters for their important roles in my personal development, their friendship, and their constant demonstration of what it looks like to strive for success. This degree would not have been possible without the many dedicated teachers I had throughout my schooling, from Mrs. Wolfe in first grade, to Mrs. Daily in fourth, to Mr. Dubow in middle school, to Mrs. Smythe, Mr. Abineri, and Mrs. Ford in high school. They, and many others, taught me how to write, how to think, how to learn, how to adapt, and how to develop as a human being. In college, I was lucky enough to find Eran Magen and James Gross, who introduced me to the field of social psychology and who allowed a know-nothing undergrad to attend lab meetings, ask dumb questions, and eventually develop an independent research program culminating in a Masters thesis. Without their patient guidance and the wealth of knowledge they shared with me, I would not have fallen in love with psychological research or applied for this PhD.

Too many people at Berkeley Haas have made an impact on me to name them all, but I want to thank my classmates, without whom I never would have escaped the curriculum-heavy first year of the program and who frequently helped me think through tough problems and tricky methodology. I also want to thank all of the MORS faculty who, to a person, have been incredibly present, available, and engaged. I want to thank Drew Jacoby-Senghor for allowing me to participate in his lab, which exposed me to new work and often enriched my own ideas. Thank you also being a gracious and magnanimous boss during my time as your GSI. I picked up so much knowledge about negotiating, and I appreciated your sense of humor throughout.

I need to thank my committee members as well. To Dacher Keltner: Thank you for lending your wisdom and perspective to my research. Your seminars were some of my favorite classes in my time at Haas; your enthusiasm is catching, and your teaching really grounded me within the emotion literature. To Don Moore: Science is better when conducted with appropriate rigor, and you have helped me to achieve the perfect level of confidence in my methods and conclusions. Your critical attention to detail also helped me to improve my research questions and fine-tune my methodologies for this dissertation. To Cameron Anderson: Thank you for guiding me through the first few years of this program with patience and thoughtfulness. You taught me how to conduct a proper literature review, and you taught me the valuable skill of moving on when your project reaches a dead end, even if it's the last thing you want to do. Without that push, I may still be spinning my wheels trying to prove that stress causes decreased performance in teams. To Juliana Schroeder, my long-suffering advisor: Thank you for recognizing the potential in me and pushing me to achieve it. Thank you for fostering my creativity and for being willing to embark on projects that were outside your area of expertise. Thank you for your meticulous edits and improvements to my writing. Thank you not only for driving me to be an independent thinker and worker, but for always being present and available when I needed to ask for help. I truly do not believe I would have completed this dissertation had you not set such a high bar for me and challenged me to meet it. I will forever be grateful for the time and effort you put into my education and my personal development. Thank you.

Thank you to Emily, Jenny, Iris, Jen, Alicia, Bree, my Nightlock teammates, and countless others who kept me grounded, offered advice, provided a necessary social outlet, pushed me to exercise, listened to me complain, drank boba with me, talked me through my anxieties, and generally helped me maintain my sanity throughout the difficult journey of earning a PhD.

And finally, thank you to my wife, Loryn. Though this is not the co-authorship I know you wanted, thank you for letting me bounce ideas off of you, for helping me prepare graphs, images, and presentations, for offering an outside perspective on some of my more ambitious research ideas, for cooking dinner when I was working late, for taking the dog to work when I needed to put my head down and focus, and for generally supporting and loving me during moments of calm and moments of stress. Without you, I would be lost. I love you.

CHAPTER 1:

Momentum Miscalibration:
Psychological Momentum Influences Expected Performance More Than Actual Performance

Abstract (Chapter 1)

People believe that psychological momentum is a force that accounts for success, but does the power of momentum reside mostly in performers' minds? Field data from an Ultimate Frisbee tournament ($N = 519$) and six experiments ($N = 2{,}533$) compared the effects of momentum on *actual* performance alongside players' *beliefs* about its effects. In Experiments 1a–d, individuals who completed a competitive task that became progressively easier (versus harder) perceived that they had more momentum and believed that they were more likely to win, but did not win more often. Experiment 2 used a signal detection analysis to quantify the extent of players' miscalibration; results revealed that individuals who gained momentum were more likely to incorrectly predict that they would win a competition, whereas individuals who lost momentum were more likely to incorrectly predict that they would lose. These errors occurred despite financial incentives to make accurate predictions and to win the competition, and with full knowledge of the competition task. Finally, individuals in Experiment 3 who gained momentum were more likely than those who lost momentum to bet on their future performance, yet they did not perform better. Experiment 3 further examines two possible moderators of the effect of momentum on expected performance: players' growth mindset about their skills, and whether they were an underdog in the competition. Overall, these results provide evidence for "*momentum miscalibration*:" psychological momentum influences performance expectations more than it influences performance outcomes.

“People who succeed have momentum. The more they succeed, the more they want to succeed, and the more they find a way to succeed.”—Tony Robbins

"We lost the momentum when we missed the field goal. They came back, and scored, and we never got back into it. It was all about momentum."—Dan Hinds

From hitting free throws to pitching strikes, athletes swear by the power of momentum. Commentators, coaches, fans, and players alike believe that momentum can determine the outcome of a game. Yet, the exact mechanisms by which momentum might actually influence performance are complex (for relevant theories, see Briki & Markman, 2018; Cornelius, Silva, Conroy, & Petersen, 1997; Taylor & Demick, 1994; Vallerand, Colavecchio, & Pelletier, 1988; Vancouver & Purl, 2017), and momentum’s actual effect on performance is hotly debated among academics and statisticians (e.g., Arkes, 2013; Miller & Sanjurjo, 2014). Does momentum matter as much as people think it does? The research presented in Chapter 1 examines how psychological momentum influences expected compared to actual performance.

Defining Psychological Momentum

To understand momentum’s consequences, it is first necessary to define it. Metaphors describing the experience of psychological momentum abound—momentum is the “wind at one’s back,” a “hot streak,” or “being on a roll.” These metaphors highlight how psychological momentum is a phenomenological experience, something that people feel at a specific point in time. It can have a transient quality, existing in one moment and then gone in the next. But does psychological momentum exist *only* in a person’s mind, or is it also grounded in physical reality?

Scholars appear to be split on this question even in their definitions of momentum. Iso-Ahola & Mobily (1980; see also Iso-Ahola & Dotson, 2014), for instance, define momentum as “an added or gained psychological power which changes interpersonal perceptions and influences an individual’s mental and *physical* performance” (emphasis added; p. 392). This definition assumes that momentum *directly influences* performance and performance outcomes (see also Adler, 1981; Taylor & Demick, 1994 for other definitions with this property). On the opposite end of the spectrum, others define psychological momentum as a “perception” rather than a real force. For instance, Vallerand et al. (1988) define momentum as the “perception that an actor is moving towards his/her goal” (p. 94). In this definition, the direct causal influence of momentum on performance is not mentioned, and actual movement toward a goal is not a requirement.

Taking a convergent position, Markman and Guenther (2007) propose that momentum is a psychological experience that need not affect performance but must involve movement toward a goal (e.g., a string of successes or failures). They write: “a precipitating event provides a target (e.g., an attitude object, person, or group of persons) with velocity, and additional precipitating events can increase velocity” (p. 801). Similarly, Taylor and Demick (1994) define psychological momentum as “a positive or negative change in cognition, affect, and/or physiology caused by an event or series of events” (p. 54).

Consistent with these latter definitions, here I define psychological momentum as the felt progress of moving toward or away from a goal preceded by an experienced change in upward or downward trajectory in pursuit of that goal (e.g., Carver & Scheier, 1990; Gernigon, Briki, & Eykens, 2010; Hubbard, 2015). Momentum perceptions, then, require that actors and observers first experience or observe a change in trajectory toward some goal. My definition notably lacks any reference to confidence or skill, components of efficacy that often accompany momentum but which are separable constructs. My conceptualization of momentum is also in line with lay participants' reported understanding of the phenomenon (see Briki, Den Hartigh, Hauw, & Gernigon, 2012; Markman & Guenther, 2007).

Momentum trajectories. The experience of moving toward or away from a goal—eliciting feelings of gaining or losing momentum—can derive from many sources. One common source is the experience of a dynamic string of successes or failures (Iso-Ahola & Dotson, 2014; Markman & Guenther, 2007; Shaw, Dzewaltowski, & McElroy, 1992). For instance, Silva, Cornelius and Finch (1992) found that participants in a loss-win-win condition reported feeling more momentum than did those in a win-loss-loss condition. Further, Vallerand et al. (1988) showed that participants rated players as having more momentum when they had won five out of ten tennis games in a pattern of three losses, then one win, then two losses, then four wins (i.e., "0001001111") than those who had won the same number of games in a pattern of one loss, one win, two losses, two wins, one loss, one win, one loss, and one win (i.e., "0100110101"). This latter manipulation of momentum contains the same number of wins and losses in both conditions, thereby manipulating only the ordering of the outcomes.

There are two possible reasons why trajectories (e.g., the ordering of wins and losses) can influence momentum perceptions. First, performers' skills could be actually getting better or worse; second, there may be something incidental in the environment that elicits these perceptions. Exemplifying the latter possibility, consider the common case in which a set of tasks or competitions vary in difficulty. If a performer happens to complete tasks in order of increasing difficulty, thereby experiencing more successes or "wins" at the beginning of the trajectory and failures or "losses" at the end, they might feel like they are losing momentum. But completing tasks in the opposite order (i.e., decreasing-difficulty) may create the perception of gaining momentum.[1]

In such cases, when performers experience gaining or losing momentum due to the difficulty order of a string of tasks that they complete (which is a situational factor), do they nevertheless tend to overestimate the contribution of their own skills, thereby drawing a more internal attribution about the cause of their momentum perceptions than is warranted? The purpose of the current chapter is to test the possibility that momentum perceptions are miscalibrating because they enhance discrepancies between expected and actual performance.

Momentum, Expected Performance, and Actual Performance

My prediction that momentum perceptions may miscalibrate performance expectations by influencing expectations more than actual performance expands upon prior research concerning how momentum and momentum-like experiences (e.g., confidence, optimism) can influence both expected and actual performance.

Expected performance. There is substantial reason to believe that experiencing momentum changes individuals' beliefs about their future performance. I categorize prior findings into three buckets that provide increasingly compelling evidence for the possibility that momentum causally affects performers' beliefs about their skills. First, correlational data demonstrate that when observers or performers experience a win (vs. a loss), the experience positively influences expectations about the next win (or loss) (Cornelius et al., 1997; Gilovich, Vallone, & Tversky, 1985; Iso-Ahola & Blanchard, 1986). Similarly, if competitors are close in

---

[1] Of course, the ordering of tasks may sometimes influence actual skill development; on novel tasks, more learning may occur when tasks are completed in an increasing-difficulty (vs. decreasing-difficulty) order (Feltz & Weiss, 1982; Robinson, 2001). In my view, however, the ordering of short sequences of familiar tasks is unlikely to meaningfully influence skill levels, and should instead only create the phenomenological *experience* of momentum.

score but one competitor has scored more points in the last few minutes or won more games in the recent past than the other, observers often expect the competitor with (perceived) momentum to win the current contest or the following game (Markman & Guenther, 2007; Vallerand et al., 1988). Moreover, individuals may act on such beliefs. For example, after basketball players make multiple successful (vs. unsuccessful) baskets, they tend to take more shots and coaches tend to substitute them out less often (Attali, 2013), and gamblers are more likely to make a bet after a win than after a loss (Croson & Sundali, 2005). However, in these data it is not clear whether the mere experience of momentum per se drives beliefs, or whether such beliefs reflect actual differences in skill levels.

Second, after manipulating an individual's or team's success or failure trajectory for the purpose of eliciting momentum perceptions, researchers have found that observers expect commensurately better or worse future performance (Briki, Den Hartigh, Markman, & Gernigon, 2014; Cornelius et al., 1997; Gilovich et al., 1985). In these experiments, however, observers may reasonably make inferences about performers' skills as a function of the information with which they are presented. Hence, inferences about skill may drive expectations more than inferences about momentum.

Third, only a few experiments (to my knowledge) have directly manipulated momentum or momentum-like experiences and subsequently measured performance expectations in performers themselves as opposed to observers (Chau & Phillips, 1995; Tenney, Logg, & Moore, 2015; Vallerand et al., 1988). Some of these experiments are inconclusive due to small sample sizes (Chau and Phillips ran 6 participants per condition and Vallerand et al. ran 11 participants per condition). Most compelling, in a statistically well-powered set of experiments, Tenney et al. (2015) demonstrated that the momentum-like experience of optimistic belief causally influences performance expectations.

Considered in aggregate, these findings indicate that momentum can causally affect beliefs about future performance, even among those who are experiencing it (as opposed to those who are merely observing it). What is less clear, however, is how such beliefs compare to actual performance. Perhaps performers are correct—or at least partly correct—in their assessments that momentum has the power to enhance (or diminish) their future performance. Alternatively, momentum perceptions may influence cognitions about performance more than they enhance performance itself.

Actual performance. Whereas perceptions of gaining momentum appear to consistently enhance expected performance, evidence for the influence of such perceptions on actual performance is decidedly mixed. On the one hand, people (and animals) that are closer to completing their goals tend to work more quickly (i.e., the "goal gradient effect"; Hull, 1932; Kivetz, Urminsky, & Zheng, 2006). Furthermore, purported evidence for the "hot hand" effect—a phenomenon whereby a performer who experiences a successful outcome has a subsequently greater chance of achieving additional success—has emerged in domains such as racquetball (Iso-Ahola & Mobily, 1980), tennis (Jackson & Mosurski, 1997; Ransom & Weinberg, 1985; Silva, Hardy, & Crace, 1988; Weinberg, Richardson, & Jackson, 1981), volleyball (Raab, Gula, & Gigerenzer, 2012; but see Miller & Weinberg, 1991), bowling (Dorsey-Palmateer & Smith, 2004; Yaari & David, 2012), and basketball (Forthofer, 1991; Larkey, Smith, & Kadane, 1989; Mace, Lalli, Shea, & Nevin, 1992; Yaari & Eisenmann, 2011), even when controls for players' actual skill levels are included. However, the effects that are reported in these studies are statistically small and the credibility of the findings has been debated within the academic community (e.g., see Arkes, 2013; Gilovich et al., 1985; Miller & Sanjurjo, 2014).

On the other hand, there are studies that report the opposite effect. For example, batting averages in baseball tend to regress toward players' means (Albright, 1993; Schall & Smith, 2000), and some experiments that manipulated momentum found that positive momentum reduced effort (e.g., among cyclists, Perreault, Vallerand, Montgomery, & Provencher, 1998; and rowers, Den Hartigh, Gernigon, Van Yperen, Marin, & Van Geert, 2014). One explanation for these findings could be that competitors tend to slack off when they feel like they are far enough ahead or have accomplished enough (Carver & Scheier, 2012; Koo & Fishbach, 2014; Louro, Pieters, & Zeelenberg, 2007). In all, evidence for the causal effect of momentum perceptions on performance is mixed at best.

Momentum Miscalibration

Even if momentum does shape both performance expectations and actual performance levels, the aforementioned literature review suggests there may be a discrepancy between the magnitudes of these effects. The influence of momentum perceptions on expectations seems to be fairly large and robust, whereas the influence of momentum perceptions on actual performance appears to be more tenuous. Comparing effect sizes across studies is problematic for many reasons, but a handful of studies have examined expectations and performance at the same time.

In one classic study (Gilovich et al., 1985), 26 basketball players believed that there would be a positive correlation between their prior free-throw successes and their next free-throw shot ($r = 0.40$), when in reality they were uncorrelated ($r = 0.02$). In another study, participants who competed in multiple rounds of a basketball shot competition against a trained confederate who shot poorly (versus shot well) became increasingly confident about winning the next round even though they actually performed no better (or slightly worse; Shaw et al., 1992). Employing a clever paradigm, Perreault et al. (1998) found that participants who fell behind a (virtual) competitor while cycling believed that they had less momentum, but they actually increased their effort when they fell behind. Finally, and perhaps most relevant to my hypothesis, Tenney et al. (2015) examined discrepancies between the effects of holding optimistic beliefs (i.e., believing in the best possible outcome) on expected and actual performance. They concluded that people think "that optimism will contribute to performance, and that sometimes this belief is wrong" (p. 393). Their findings partly support my momentum miscalibration hypothesis in the sense that optimism and momentum are related constructs, but a primary difference is that optimism beliefs need not necessarily coincide with perceptions of goal progress, whereas the phenomenological experience of momentum corresponds directly to goal progress perceptions.

The momentum miscalibration hypothesis. Prior findings lay the conceptual groundwork for my hypothesis that experiencing momentum can produce discrepancies between expected and actual performance. However, no set of experiments has cleanly tested this hypothesis.

To more fully understand how momentum perceptions can be miscalibrating, I draw on signal detection theory (Nevin, 1969) for evaluating a dichotomous outcome (e.g., winning a competition). This theory considers four categories of results: expecting to win and actually winning (hit), expecting to lose and actually losing (correct rejection), expecting to win but actually losing (miss), and expecting to lose but actually winning (false alarm); the first two categories are accurate predictions whereas the latter two are inaccurate. My *momentum miscalibration* hypothesis suggests that momentum perceptions (as compared to the lack thereof) will in general lead people to make relatively more inaccurate predictions, and that the nature of these inaccurate predictions will depend upon whether people perceive that they are gaining versus losing momentum. Specifically, perceptions of gaining momentum will increase the likelihood of misses whereas perceptions of losing momentum will increase the likelihood of false alarms.

Consequences. If momentum miscalibration occurs, it may have potentially problematic downstream consequences. For one, people's inflated beliefs about their own skills and abilities could lead them to act unwisely. For example, they may be more likely to start a task that they are unlikely to complete, or they may be more likely to bet on their own performance. In the United States, sports betting, both illegal and legal, is estimated to be a $500 billion a year industry, and legal gambling in the United Kingdom generates over $7 billion a year in annual GDP (Polisano, 2018). If the betting behavior of individuals is at least partly based on an exaggerated, if not incorrect, assumption that (perceived) momentum causally influences performance, their betting tendencies may have significantly negative financial consequences. Indeed, momentum perceptions are correlated with betting behavior in sports markets such as the NBA and NFL (Arkes, 2011; Avery & Chevalier, 1999; Hartzmark & Solomon, 2012; Williams, 2010), in the stock market (Dunn, 2000; Jegadeesh & Titman, 1993), and in gambling (Rabin & Vayanos, 2010; Sundali & Croson, 2006; Tversky & Kahneman, 1971). Thus, momentum miscalibration may encourage betting behavior that misaligns with actual performance outcomes.

Overview of Studies

The current chapter seeks to provide empirically robust and replicable tests of the momentum miscalibration hypothesis. I present six experiments that utilize a clean and replicable manipulation of momentum: changing the order by which performers complete practice tasks (increasing-difficulty or decreasing-difficulty order) prior to a competition task. In each experiment, performers complete practice tasks in one order or the other, report their experienced momentum, and then make predictions about their competition performance (which I also measure). I selected a discrete task that all participants could conceivably complete and for which I could empirically vary the difficulty level: a word search task.

Experiments 1a–d compare the effect of completing tasks in increasing-difficulty order versus decreasing-difficulty order on perceived momentum, expected performance, and actual performance on a competition task. I conducted four different versions of this experiment both for internal replication purposes (to ensure the effects are robust) and to examine variations in the final performance task, seeking the most conservative possible test of my miscalibration hypothesis. In other words, I deliberately tested conditions that I thought would be optimal to create an effect of momentum on real performance, if such an effect exists.

Experiment 2 provides a more precise measure of miscalibration by examining discrepancies between performance predictions and performance levels among participants who are either gaining or losing momentum. Importantly, I maximize both the likelihood that participants' predictions are informed (because participants are familiar with the competition task and have seen it in advance) and accurate (by using financial incentives for accuracy). I also financially incentivize real performance to ensure participants are trying to win. This paradigm allows me to test whether experimental condition influences the likelihood of making an inaccurate estimate of one's own performance (i.e., a false alarm or miss) versus an accurate estimate (i.e., a correct rejection or hit).

Finally, Experiment 3 tests a potential consequence of being miscalibrated. I give participants an opportunity to bet on their performance, predicting that those assigned to complete their tasks in decreasing-difficulty versus increasing-difficulty order will be more likely to bet on themselves despite being no more likely to actually win. To be thorough, Experiment 3 also examines two possible moderators of the effect of momentum on expectations: performers' favorite or underdog status, and their growth mindset about their performance. Performers' underdog status could influence momentum miscalibration because, at least under certain circumstances, fans prefer underdogs to favorites, even believing they will win despite their lower status (Ceci & Kain, 1982; Frazier & Snyder, 1991; Vandello, Goldschmied, & Richards, 2007). It may therefore be possible that the presence of momentum could increase expected (vs. actual) success particularly for underdogs. Performers' mindset could influence momentum miscalibration because growth-oriented mindsets are associated with improved performance (at least in academic domains; e.g., Dweck, 2000; Paunesku et al., 2015). Experiment 3 therefore tests whether any interactions exist between performers' status, their growth mindset, and their experienced momentum on their expected and actual performance.

Across experiments, I also seek to understand performers' attributions of their momentum. In all studies, performers report whether their performance was getting better or worse, and their resulting self-efficacy levels (i.e., a more internal attribution for their momentum). Experiment 3 further measures whether performers knew that the task difficulty ordering was changing (i.e., an external attribution). I expect that, despite knowing that the task difficulty ordering was changing, performers would tend to internally attribute the momentum to their own skill, which may in part drive the momentum miscalibration result.

For all studies, I report how I determined my sample size, all data exclusions, and all measures (Simmons, Nelson, & Simonsohn, 2011). All data, code, and survey materials are available in the Open Science Framework repository for this project (https://osf.io/qqng5/?view_only=04e2e3ec43b747c7a86754ca0d0482a8).

Field Data from an Ultimate Frisbee Tournament

Before conducting experiments, I first examined my hypothesis that perceived momentum will be more strongly associated with performance expectations than it will be with performance outcomes in a competitive, real-world setting. To do so, I collected 519 survey responses from 107 frisbee players (93 female, M_{age} = 27.40 years, 95% CI [26.36, 28.44]) who played 42 games at the USA Ultimate 2017 Southwest Club Regionals Ultimate Frisbee Tournament. Players completed a survey before (N = 204) and after (N = 315) every game at the tournament; some players completed multiple surveys but not all players completed every survey.

Before each game, players reported their pre-game momentum ("How much momentum does your team have, entering this game?" -50 = *negative momentum*, 50 = *positive momentum*) and their confidence about winning ("I am confident about this game," 1 = *strongly agree*, 7 = *strongly disagree*; reverse-coded for analysis). A subset of these players ($N_{Pre\text{-}game}$ = 179, $N_{Post\text{-}game}$ = 266) also completed an intake survey in which they reported their tenure ("How many years have you been on the team?"), the number of hours they practice per week, and their enjoyment with being on the team ("Are you pleased with your decision to play on this team this season?" 1 = *Yes, I love this team*, 5 = *No, I wish I hadn't played*). I also collected each team's ranking before the tournament to control for skill level.

Because this data is nested, with multiple responses from single individuals, individuals nested within teams and ratings nested within games, I ran a cross-classified multilevel model to analyze the results. Level 1 units were ratings of momentum (n = 204), with individuals as Level 2 units (n = 85) and teams as Level 3 units (n= 8). Ratings are also nested within games (n = 33; Level 2 unit). Ratings were person-mean centered. To test whether ratings of momentum predicted confidence in winning, I ran a model with person-mean ratings of momentum before the game predicting confidence in winning, including random intercepts for games and for individuals nested within teams.

Consistent with my prediction, pre-game momentum perceptions and winning confidence were positively correlated, $\beta = 0.03$, 95% CI [0.01, 0.04]. This relationship remained statistically significant after controlling for tenure, practice hours, enjoyment, and ranking, $\beta = 0.02$, 95% CI [0.01, 0.04].

I next examined whether pre-game momentum perceptions correlated with actual performance levels. To control for teams' prior performance, I computed a performance score by using an algorithm created and maintained by USA Ultimate[2] that takes into account every game played during the season prior to the regional tournament and predicts the number of goals the losing team should score in each game. I subtracted the number of goals actually scored from the number of predicted goals to compute a relative performance score ($M = -0.52$, 95% CI [-1.05, 0.02]). This metric is more informative than a binary win-loss measure because it accounts for teams' skill. Pre-game momentum perceptions did not correlate with performance scores, $\beta = 0.05$, 95% CI [-0.01, 0.11]. This null result remained unchanged after controlling for tenure, practice hours, enjoyment, and ranking in a simple linear model, $\beta = 0.06$, 95% CI [0.00, 0.13].

Finally, I examined whether post-game momentum perceptions ("How much momentum does your team have, after playing that game?" -50 = *negative momentum*, 50 = *positive momentum*) correlated with performance scores. I ran the same model as above, substituting post-game momentum ratings for pre-game momentum ratings. The two variables were correlated, $\beta = 0.12$, 95% CI [0.10, 0.13], suggesting that performers may infer momentum perceptions from prior performance outcomes, and the relationship remained statistically significant after controlling for age, gender, tenure, practice hours, enjoyment, and ranking, $\beta = 0.12$, 95% CI [0.10, 0.14]. Overall, these field data provide preliminary evidence to suggest that momentum perceptions better predict performance expectations than performance outcomes.

Experiments 1a–d: Momentum, Expected Performance, and Actual Performance

To examine how momentum perceptions causally influence expected and actual performance levels on a competition task, participants in Experiments 1a–d completed three practice rounds of a task in increasing-difficulty or decreasing-difficulty order, which I expected to generate feelings of losing or gaining momentum, respectively, and then engaged in a competition.

Common methods across Experiments 1a–d

Experiments 1a–d each employed two experimental conditions (between-subjects): *increasing-difficulty order* and *decreasing-difficulty order*. After signing a consent form and before beginning the word search task, participants read that they would "see twelve letters" and then have "one and a half minutes to write down as many 4+ letter words as possible using those letters." The word-find task had four rules: 1) each word could only be submitted once; 2) each letter could only be used once per word; 3) each word had to be at least four letters long; and 4) the words had to be real words that could be found in a dictionary. Participants' goal was to find as many words as possible following the four rules in the time allotted for each practice round and the competition. For example, if their letter string was "*XHWYNEAJRTMF*," they could write words such as: "*near," "meat," "fret," "wart," "wharf,"* and so on. I told participants that they would be completing three practice rounds of the task before the competition, and they

[2] https://play.usaultimate.org/teams/events/rankings/#algorithm

responded to several attention check items to ensure that they had read and understood the rules (see Appendix 1).

Manipulating task-difficulty. To classify the difficulty level of the word-find tasks, I pretested six letter sets to determine the average number of words found in 1.5 minutes. The pretest results are reported in Appendix 1. I selected three letter sets: easy ("AEITFMNLPRYG"; $M_{\text{words found}}$ = 15.10, 95% CI [13.00, 17.19]), medium ("AEIBTJNCKYDH"; $M_{\text{words found}}$ = 10.60, 95% CI [9.09, 12.11]), and hard ("EIOBTJNCMYRP"; $M_{\text{words found}}$ = 8.08, 95% CI [6.76, 9.40]). In the *decreasing-difficulty* condition, participants completed the hard letter set, then the medium letter set, and then the easy letter set. In the *increasing-difficulty* condition, conversely, they completed the same letter sets in the opposite order—easy, then medium, then hard. Importantly, both conditions contained the same tasks, therefore giving participants the same set of experiences in aggregate. To confirm that participants' skill levels did not change more in one condition than the other, I computed the actual scores for each practice round.

Awareness of changing performance. To measure whether participants noticed their changing performance across rounds, I asked, "How do you think your performance is changing over time?" (-3 = *I'm getting a lot worse*, 3 = *I'm getting a lot better*) prior to Practice Round 3 and prior to the competition.

Perceived momentum. To measure momentum perceptions, I asked, "How much **momentum** do you think you have right now, headed into [Practice Round 3]?" To respond, participants moved a slider that was anchored at 0 (*No momentum*) and 100 (*A lot of momentum*), and was initially set at 0. Perceived momentum was measured three times: prior to Practice Round 3, prior to the competition, and after the competition.[3]

Perceived self-efficacy. Although my primary prediction was that the difficulty ordering of tasks would influence perceived momentum and, consequently, performance expectations, I further speculated that participants' beliefs about their momentum might also influence self-efficacy levels, as self-efficacy has long been linked to performance outcomes (Bandura, 1977; Shaw, Dzewaltowski, & McElroy, 1992). To assess self-efficacy I asked three questions in Experiments 1a-c—"How skilled do you think you will be at finding words?", "How confident do you feel about finding words?", and "How much do you trust your ability to find words?" (1 = *not at all* [*confident/skilled/much*], 7 = *very* [*confident/skilled/much*]; $\alpha s \geq .97$)—at five time-points throughout the experiment: before the first practice round, after each practice round, and after the competition.

Expected performance measures. After completing Practice Round 3 and responding to the manipulation check, momentum, and self-efficacy items, participants learned about their opponent for the competition round: "The opponent has already been selected. Your opponent has consistently found **X** words in their practice rounds. In order to win, you will probably need to find more than **X** words," where **X** was participants' own average score from practice rounds 1-3 rounded to the nearest integer.[4] I used participants' average score in the description of their (ostensible) opponent for two reasons: first, using this score made it seem feasible that participants could either win or lose, thereby creating some potential for variability in

[3] I did not ask this question in earlier practice rounds because the initial perception of momentum also requires a perceived change in trajectory, and thus I reasoned that momentum perceptions could only be measured once participants had completed at least two rounds of word-find tasks. I expected that momentum perceptions would change as a function of momentum condition, but I made no *a priori* predictions regarding how momentum perceptions might change after the competition was over.

[4] I changed the framing of the competition slightly for Experiment 1d; see Experiment 1d Method section.

expectations, and second, it standardized participants' beliefs about their opponent across conditions, ensuring that different beliefs about the competition or one's opposition could not be driving my results. I also reminded participants of their own scores in each practice round. I then asked two measures of expected performance just after Practice Round 3 and prior to the competition: 1) "How likely are you to win the competition?" (1 = *not at all likely*, 10 = *very likely*) and 2) "Who do you expect to win the competition?" (*You*, *Your opponent*).

Actual performance measures. Participants then completed the competition, which was a novel set of letters that I selected from my pretest ("AEODTSNCPYRK"; $M_{\text{words found}}$ = 14.98, 95% CI [12.38, 17.58]). For reasons described in the separate Method sections below, I varied the timing of the competition in each experiment. The competition was untimed in Experiments 1a and 1b, 3 minutes long in Experiment 1c, and 1.5 minutes long in Experiment 1d. Participants' competition score (number of words found) was the primary measure of actual performance.

Post-competition feelings. Following the competition, participants completed several exploratory measures. In Experiments 1a–c, in addition to recording their self-efficacy levels for the fifth time and their momentum perceptions for the third time, they reported their performance feelings ("Overall, how did you feel about your performance in the competition?" (1 = *not at all good*, 10 = *excellent*)) and competition effort ("How much effort did you put into the competition?" (1 = *no effort*, 10 = *a lot of effort*)). In Experiment 1d, I included two more measures to examine alternative possibilities for how experimental condition might affect performance expectations: "How fun did you think this game was?" (*Not at all fun*, *A little fun*, *Somewhat fun*, *Very fun*, *Extremely fun*) and How are you feeling right now?" (*Extremely unhappy*, *Moderately unhappy*, *Slightly unhappy*, *Neither happy nor unhappy*, *Slightly happy*, *Moderately happy*, *Extremely happy*).

Control variables. To control for participants' experience with the task, in all experiments I asked about task familiarity: "How familiar are you with word-find tasks similar to the ones you completed today?" (*I have never played a game like that before*, *I have played a game like that a few times*, *I sometimes play games like that*, *I frequently play games like that*, *I play games like that almost every day*).

Methods unique to each experiment

Experiment 1a method. Based on three prior experiments that tested the effect of a different momentum manipulation on performance beliefs among observers (*d*s = 1.00, 0.41, and 1.13; see Appendix 1 for details), I aimed for a sample size of 80 participants in each of the two experimental conditions.[5] I recruited 159 participants (87 female, 1 gender non-binary, M_{age} = 35.52 years, 95% CI [33.66, 37.37]) through Amazon Mechanical Turk who completed a survey in exchange for $1.00. This experiment followed the procedure described above.

Experiment 1b method. In addition to testing my primary hypothesis, Experiment 1b also examined whether interrupting perceived momentum would diminish its effect on expectations. Consistent with Experiment 1a, I predetermined a sample size of 80 participants that were randomly assigned to each of the conditions in a 2 (task order: increasing-difficulty or decreasing-difficulty) × 2 (interruption: no interruption or interruption) between-participants design. In total, I recruited 339 participants (161 female, 5 gender non-binary, M_{age} = 34.04 years, 95% CI [33.00, 35.08]) through Amazon Mechanical Turk who completed the survey in

[5] In retrospect, it may not have been ideal to use these experiments for my power analysis, since they relied on a substantially different paradigm. I increased the sample size in Experiments 1c and 1d to account for this possibility.

exchange for $1.00. To interrupt momentum, I randomly assigned half of my participants to see the following message prior to the competition task for one minute: "ERROR! QUALTRICS IS EXPERIENCING AN ERROR. PLEASE WAIT WHILE WE FIX THIS ERROR," along with a moving image of a loading circle. After the page re-loaded, I measured perceived momentum and self-efficacy again to see whether they had been affected by the interruption manipulation. Participants then completed the competition task and survey items described in the Common Methods section.

Experiment 1c method. The purpose of Experiment 1c was to run a high-powered replication test of Experiment 1a (pre-registered at https://osf.io/bsdzu), and I therefore predetermined to recruit about five times as many participants (i.e., 350/condition). I was concerned that my previous experiments might have been under-powered to detect actual performance differences. In total, through Amazon Mechanical Turk I recruited even more participants than expected due to a glitch in the survey that created a re-start (N = 917 participants; 491 female, 8 gender non-binary, M_{age} = 36.18 years, 95% CI [35.42, 36.95]). All participants received $1.00 in compensation. This experiment followed the same procedure as Experiment 1a with the exception that the competition task was timed for 3 minutes (instead of being allotted unlimited time). I made this change to examine whether time pressure would alter the effect of perceived momentum on performance.

Experiment 1d method. Experiment 1d examined whether perceived momentum would influence participants' willingness to compete (pre-registered at https://osf.io/7gvp6). Based on the effect sizes obtained in Experiments 1a–c, I predetermined a sample size of 200 participants in each of two experimental conditions. I recruited 407 participants in total (236 female, 1 gender non-binary, M_{age} = 37.22 years, 95% CI [36.03, 38.42]) through Amazon Mechanical Turk who completed a survey in exchange for $0.75. The procedure was the same as Experiment 1a through the first three practice rounds, with the exception that I omitted the self-efficacy questions in order to shorten the survey. In this experiment, however, I made the competition round optional and explicitly told participants that they would be competing against a number and not a real opponent (to reduce deception). I told them, "Now that you are done with the first three rounds, you have the opportunity to play one more round if you would like to. In this round you will compete against the average of everyone who has played this set of letters before. The average for this set of letters is **X**. In order to win you will need to find more than **X** words in 1.5 minutes," where **X** was their own average score from the first three rounds. To increase the time pressure even more from Experiment 1c (3 minutes), I reduced the competition time to 1.5 minutes in Experiment 1d. This amount of time also matched the practice rounds, making the competition more similar to the practice rounds.

After providing participants with the competition information, I asked, "Do you want to enter the competition? If you choose yes, you will play the next round for 1.5 minutes, then you will find out your score (whether you won or lost), answer a few questions, and finish this experiment. If you choose no, you will not play the next round, and instead just answer a few questions and finish the experiment. Please make your selection below:" (*Yes*, *No*). Those participants who chose to compete learned their scores and whether they had won or lost after the competition task was over. All participants then saw the same set of final questions employed in Experiments 1a–c.

Results

Analysis strategy. In all experiments, I removed participants from analyses who did not perform within one standard deviation below the mean on *either* their combined practice round average score or their competition score. I did this because I determined that participants who scored under these thresholds were either not trying very hard or were so deficient at the task that their results would not accurately reflect the effect of task-difficulty condition. Including these participants in analyses does not significantly change any of my results, as described in Appendix 1.

The results on the variables of interest in each experiment were so similar that I collapsed my analyses across all four experiments. In Experiment 1b, my interruption manipulation had no effect on any of my variables of interest, so I collapsed these conditions into *decreasing-difficulty* and *increasing-difficulty* for analysis.[6] In Experiment 1d, 71.5% of participants decided to compete, and I only included these participants in my analyses of actual performance. In total, then, I included 1,386 participants in the analyses across all four experiments prior to the competition (701 in the *decreasing-difficulty* condition, and 685 in the *increasing-difficulty* condition). I also included a dummy variable in the analyses to control for the fixed effects of each experiment.

Practice round scores. First, actual practice round scores were not meaningfully different between the increasing (M =11.45, 95% CI [11.13, 11.76]) and decreasing difficulty (M = 11.88, 95% CI [11.53, 12.22]) conditions, $t(1384) = 1.80$, $p = .072$, $d = 0.10$, suggesting that my manipulation did not strongly affect actual skill levels.

Awareness of changing performance. Participants indeed noticed that their performance was changing across rounds; they reported improving more on the practice rounds in the *decreasing-difficulty* condition (M = 1.05, 95% CI [0.95, 1.14]) than in the *increasing-difficulty* condition (M = -0.47, 95% CI [-0.57, -0.36]), $t(1384) = 21.44$, $p < .001$, $d = 1.15$.

Perceived momentum. Supporting my first prediction, participants in the *decreasing-difficulty* condition perceived that they had significantly more momentum immediately prior to the competition round (M = 63.9, 95% CI [62.24, 65.57]) than did participants in the *increasing-difficulty* condition (M = 44.7, 95% CI [42.76, 46.58]), $t(1384) = 14.94$, $p < .001$, $d = 0.80$. Perceived momentum after Practice Round 2 followed a similar pattern but did not show as large a difference as perceived momentum just before the Competition Round ($M_{\text{decreasing}} = 54.37$, 95% CI [52.67, 56.08]; $M_{\text{increasing}} = 51.95$, 95% CI [50.15, 53.75]), $t(1384) = 1.92$, $p = .055$, $d = 0.10$, likely because participants had not acquired the amount of experience with the task that would be sufficient to establish a sense of momentum. Self-efficacy followed a similar pattern across practice rounds (see Appendix 1).

Expected performance. Supporting my second prediction, task-difficulty condition had a statistically significant effect on participants' performance expectations, $t(1384) = 4.70$, $p < .001$, $d = 0.25$, as participants in the *decreasing-difficulty* condition believed that they were more likely to win the competition (on the 1 to 10 Likert scale item; M = 5.61, 95% CI [5.44, 5.78]) than were participants in the *increasing-difficulty* condition (M = 5.03, 95% CI [4.86, 5.20]). This effect also emerged in each experiment separately except for Experiment 1a (1a: $p = .813$, 1b: $p = .039$, 1c: $p = .003$, 1d: $p < .001$). In a regression analysis that predicted the influence of condition on expectations, the effect remained statistically significant after controlling for

[6] 18 participants dropped out of the experiment after encountering the interruption, but I continued to collect data from participants until I had a relatively equal number in each condition and achieved my predetermined sample size of 320 participants. I only include participants in this analysis who completed the full experiment.

average practice round score, age, gender, education, and task familiarity (see Models 1 and 2 in Table 1).[7]

Participants in the *decreasing-difficulty* condition were also more likely to expect that they were going to win the competition (on the dichotomous measure; win = 1, lose = 0) (M = 0.61, 95% CI [0.57, 0.65]) than were participants in the *increasing-difficulty* condition (M = 0.53, 95% CI [0.49, 0.57]), $t(1147) = 2.59$, $p = .010$, $d = 0.14$. However, I note that this effect was relatively weak across studies and should be interpreted with caution (1a: $p = .284$, 1b: $p = .046$, 1c: $p = .120$). The aggregated effect remained robust after controlling for baseline self-efficacy levels, average practice round score, age, gender, education, and task familiarity.

Table 1. *Summary of Hierarchical Regression Analysis for Variables Predicting Perceived Likelihood to Win Controlling for Experiment Number (in Experiments 1a–d)*

	Model 1			Model 2			Model 3			Model 4		
Variable	*B*	*SE B*	*t*	*B*	*SE B*	*t*	*B*	*SE B*	*t*	*B*	*SE B*	*t*
Experimental condition	0.54	0.12	4.46***	0.54	0.11	4.76***						
Average practice score	0.11	0.01	7.82***	0.07	0.01	5.40***	0.03	0.01	2.20*	0.02	0.01	1.57
Age				-0.01	0.01	-1.53				-0.02	0.01	-2.63**
Gender				-0.05	0.08	-0.56				-0.05	0.07	-0.75
Education				0.52	0.04	13.91***				0.28	0.03	8.16***
Familiarity with task				0.10	0.07	1.40				0.06	0.07	0.88
Perceived momentum							0.00	0.00	0.42	0.00	0.00	0.97
Perceived self-efficacy							0.60	0.04	15.59***	0.50	0.04	12.77***
Baseline self-efficacy							0.00	0.01	-0.57	0.00	0.01	-0.87
Adjusted R^2	0.06			0.18			0.39			0.43		
N	1385			1385			1148			1148		

Note: Models 1 and 2 include Experiment 1d but Models 3 and 4 do not because I did not measure self-efficacy in Experiment 1d. Perceived momentum and self-efficacy were measured immediately prior to the competition round. Baseline self-efficacy was measured before the first practice round. Experimental condition is coded as 1 = decreasing difficulty order; 0 = increasing difficulty order. *$p < .05$. **$p < .01$. ***$p < .001$

[7] Interestingly, albeit unpredicted, removing experimental condition from the regression analysis revealed only a significant relationship between self-efficacy levels and expectations (see Models 3 and 4 in Table 1), indicating that participants' self-efficacy levels just before the competition positively correlated with their performance expectations even after controlling for momentum perceptions before the competition, baseline self-efficacy levels, practice round score, and demographics.

Momentum mediates performance expectations. A 10,000 sample bootstrapping mediation model indicated that the effect of task-difficulty condition on expectations was mediated by momentum perceptions (95% bias corrected CI = [0.79, 1.05]).[8]

Actual performance. I next examined the effect of task-difficulty condition on competition task performance. Despite the fact that task-difficulty condition enhanced performance expectations, condition neither affected performance levels, $t(1384) = 1.06$, $p = .289$, $d = 0.06$ ($M_{\text{decreasing}} = 24.75$, 95% CI [23.89, 25.61]; $M_{\text{increasing}} = 24.10$, 95% CI [23.26, 24.94]), nor the actual likelihood of winning, χ^2 (1, $N = 1386$) = 1.23, $p = .267$, and these null effects remained unchanged after controlling for average practice round score, age, gender, education, and task familiarity (see Models 1 and 2 in Table 2).

I further examined the effect of task-difficulty condition on competition task performance within each separate experiment because I changed the competition rules in each experiment. In the untimed competition task in Experiment 1a, participants in the *decreasing-difficulty* condition found marginally more words ($M = 35.55$, 95% CI [32.39, 38.70]) than did those in the *increasing-difficulty* condition, ($M = 31.41$, 95% CI [28.09, 34.73]), $t(125) = 1.80$, $p = .074$, $d = 0.32$, although they spent the same amount of time on the competition task ($M_{\text{decreasing}} = 317.92$ sec, 95% CI [275.66, 360.18]; $M_{\text{increasing}} = 345.33$ sec, 95% CI [287.17, 403.49]), $t(125) = 0.77$, $p = .443$, $d = 0.14$. In Experiment 1b, which also used an untimed competition task, participants found *fewer* words in the *decreasing-difficulty* condition ($M = 26.39$, 95% CI [24.24, 28.55]) than they found in the *increasing-difficulty* condition ($M = 29.01$, 95% CI [26.60, 31.41]), $t(269) = -1.60$, $p = .110$, $d = -0.20$, and participants spent the same amount of time on the competition task ($M_{\text{decreasing}} = 277.72$ sec, 95% CI [240.51, 313.75]; $M_{\text{increasing}} = 277.13$ sec, 95% CI [245.61, 309.83]), $t(269) = 0.02$, $p = .981$, $d < 0.01$. In Experiment 1c, during which participants had three minutes to complete the competition task, I found no difference in performance scores between the *decreasing-difficulty* ($M = 25.35$, 95% CI [24.34, 26.36]) and *increasing-difficulty* conditions ($M = 24.30$, 95% CI [23.42, 25.18]), $t(749) = 1.54$, $p = .125$, $d = 0.11$. In Experiment 1d, during which participants had only 1.5 minutes to complete the competition task, I once again found no difference between performance scores in the *decreasing-difficulty* ($M = 14.85$, 95% CI [13.94, 15.75]) and *increasing-difficulty* conditions ($M = 14.23$, 95% CI [13.18, 15.27]), $t(235) = 0.89$, $p = .375$, $d = 0.12$. I also tested whether condition affected participants' desire to enter the competition, but it did not, $t(405) = 0.66$, $p = .511$, $d = 0.07$ (the competition task in Experiment 1d was optional, and thus only those participants who chose to compete (71.5%) were included in the analyses reported above).

Because I observed no effect of experimental condition on performance, I further conducted an exploratory analysis in which I tested correlational predictors of competition task performance in regression models (see Table 2). This revealed a significant relationship between momentum perceptions and performance levels that persisted after including all of the control variables (see Models 3 and 4 in Table 2). In other words, participants' momentum perceptions predicted their performance levels even after accounting for prior performance and self-efficacy. Interestingly, the relationship between perceived momentum and performance was meaningfully larger than the relationship between self-efficacy and performance, suggesting, perhaps, that momentum perceptions more closely align with performance outcomes than do self-efficacy levels.

---

[8] To be thorough, I also tested whether self-efficacy mediated the effect, and it did: 95% bias corrected CI = [0.08, 0.28]. I also found support for a two-step model indicating that momentum perceptions and self-efficacy levels sequentially mediated the effect of task-difficulty condition on expectations, 95% bias corrected CI = [0.67, 0.95].

Table 2. *Summary of Hierarchical Regression Analysis for Variables Predicting Competition Performance Controlling for Experiment Number*

	Model 1			Model 2			Model 3			Model 4		
Variable	*B*	*SE B*	*t*	*B*	*SE B*	*t*	*B*	*SE B*	*t*	*B*	*SE B*	*t*
Experimental condition	-0.09	0.42	-0.21	-0.15	0.42	-0.35						
Average practice score	1.47	0.05	31.29***	1.42	0.05	28.79***	1.48	0.06	26.05***	1.43	0.06	24.42***
Age				0.04	0.03	1.54				0.10	0.04	2.56*
Gender				0.91	0.31	2.96**				0.95	0.35	2.72**
Education				0.07	0.14	0.51				-0.07	0.16	-0.40
Familiarity with task				0.45	0.27	1.66				0.63	0.33	1.93
Perceived momentum							0.04	0.02	2.55*	0.04	0.02	2.71**
Perceived efficacy							0.00	0.18	-0.02	-0.05	0.19	-0.25
Baseline self-efficacy							-0.01	0.03	-0.33	-0.01	0.03	-0.55
Adjusted R^2	0.53			0.54			0.44			0.43		
N	1385			1385			1148			1148		

Note: Models 1 and 2 include Experiment 1d but Models 3 and 4 do not because I did not measure self-efficacy in Experiment 1d. Perceived momentum and self-efficacy were measured immediately prior to the competition round. Baseline self-efficacy was measured before the first practice round. Experimental condition is coded as 0 = increasing difficulty order; 1 = decreasing difficulty order. *$p < .05$. **$p < .01$. ***$p < .001$

Momentum miscalibration? To examine miscalibration, I intended to compare participants' expectation of winning (1 = expected win; 0 = expected loss) with whether or not they actually did win (1 = actual win; 0 = actual loss) in a signal detection test. However, fortunately for my participants' bonus amounts but unfortunately for my intended test, over 97% of participants in Experiments 1a–d "won" the competition because I selected a relatively easier task for the competition. Therefore I could not quantify the extent of miscalibration in these studies. I use a competition task of the same average difficulty level of the practice round tasks in Experiment 2 to provide enough variance in actual performance to test for miscalibration.

Post-competition feelings. Although I did not have *a priori* predictions, to be thorough I tested the effect of experimental condition on reported effort, competition feelings, post-competition self-efficacy and momentum, evaluation of task, and mood. In Experiments 1a–c, participants did not report expending differential levels of effort on the competition task

($M_{\text{decreasing}}$ = 8.65, 95% CI [8.52, 8.78]; $M_{\text{increasing}}$ = 8.58, 95% CI [8.44, 8.72]), $t(1141) = 0.65$, $p = .515$, $d = 0.04$. However, they did report feeling better about their performance in the *decreasing-difficulty* condition (M = 6.92, 95% CI [6.74, 7.09]) than they did in the *increasing-difficulty* condition (M = 6.66, 95% CI [6.48, 6.83]), $t(1145) = 2.04$, $p = .042$, $d = 0.12$. They also reported elevated post-competition momentum perceptions (M = 68.83, 95% CI [66.94, 70.73]) and higher self-efficacy levels (M = 7.01, 95% CI [6.83, 7.18]) in the *decreasing-difficulty* condition than in the *increasing-difficulty* condition (M_{momentum} = 63.75, 95% CI [61.81, 65.69]; M_{efficacy} = 6.75, 95% CI [6.58, 6.93]), ts > 2.03, ps < .043, ds > 0.12, despite performing equivalently well on the competition task. No statistical differences by condition emerged on participants' moods or beliefs about how fun the task was in Experiment 1d, ts < 0.19, ps > .857, ds < 0.03, and controlling for these variables did not change the null effect of task-difficulty condition on competition task performance, $t(232) = 0.92$, $p = .358$, $d = 0.12$.

Discussion

Experiments 1a–d provide evidence that merely completing a set of practice tasks in a different difficulty order influences the perception of psychological momentum, which then exerts a causal effect on performance expectations. Notably, the effect on expectations occurred despite participants displaying similar skill levels and reporting similar task experience across conditions. In support of the *momentum miscalibration hypothesis*, the effect of the manipulation on expected performance (d = 0.25) was more than four times as large as the effect on actual performance (d = 0.06). However, making this comparison is problematic because expectations were measured on a Likert scale whereas actual performance was measured based on number of words found, and I could not perform a signal detection analysis because the majority of participants won the competition task. I made several changes to the paradigm in Experiment 2 to provide a more direct test of the *momentum miscalibration hypothesis*.

Despite the lack of observable causal effect of the momentum manipulation on actual performance, momentum perceptions did appear to be at least correlationally related to the competition performance scores after controlling for observable differences in skill and perceptions of self-efficacy. I return to the important question of whether, and if so how, psychological momentum might potentially influence performance outcomes in other domains in the discussion section for this chapter.

Experiment 2: Momentum Miscalibration

Experiment 2 provides a more direct test of momentum miscalibration by collecting performers' exact predictions of their competition performance and comparing them to actual performance. To make the predictions as informed as possible, I let participants view the competition task before their predictions and selected a competition task that was representative of participants' practice round experiences (i.e., of average difficulty). To make predictions as accurate as possible, I added a monetary incentive for prediction accuracy. I also added a larger monetary incentive for performance, both a base bonus if individuals won the competition and extra pay for each additional word found beyond their target. The incentive for real performance was higher than the incentive for prediction accuracy so that participants would not be tempted to perform in line with their predictions. I planned to conduct a signal detection analysis of participants' expectations and actual likelihood of winning, expecting that there would be relatively more misses (i.e., expecting to win but losing) when performers believed they were gaining momentum but relatively more false alarms (i.e., expecting to lose but winning) when performers were losing momentum.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/wut82/).[9]

Participants. Based on the effect sizes obtained in Experiments 1a–d, I predetermined a sample size of 150 participants for each of two experimental conditions. I recruited 302 participants (174 female, 1 gender non-binary, M_{age} = 36.51 years, 95% CI [35.19, 37.83]) through Amazon Mechanical Turk who completed a survey in exchange for $0.70 with the opportunity to earn a substantial bonus.

Pretest. I recruited 53 participants on Amazon Mechanical Turk (32 female, 1 gender non-binary, M_{age} = 35.30, 95% CI [32.36, 38.24]) to pretest a new set of letters (EIOBTHNCMYRG) for the competition round that were designed to be closer to the medium-difficulty letter set that participants viewed in the practice rounds. Participants' average score on this letter set (*M* = 9.29, 95% CI [7.71, 10.89]) was similar to the pretest results of the medium-difficulty letter set used during the practice rounds (see Manipulating task difficulty section in Experiments 1a–d), *M* = 10.60, 95% CI [9.09, 12.11].

Procedure. I randomly assigned participants to one of two conditions: *decreasing-difficulty* and *increasing-difficulty*. The word-find tasks employed in the practice rounds described in Experiments 1a–d were used again, and participants saw the same instructions and responded to the same attention check items employed in Experiments 1a–d. The procedure then followed Experiment 1a nearly identically through the three practice rounds, with the only difference being that the perceived momentum question was converted to a -50 (*losing momentum*) to +50 (*gaining momentum*) scale (with the slider bar starting at 0) because I believed that these scale labels would be more intuitive for participants.

After the third practice round, I told participants: "In this round you will compete against the average of everyone who has played this set of letters before. The average for this set of letters is **X**. In order to win you will need to find more than **X** words in 1.5 minutes," where **X** was actually their own average score from the first three rounds.

Incentivized performance predictions. Importantly, participants' performance was incentivized in the following way: "If you win, you will earn a $0.30 bonus. On top of that, we will give you $0.05 for every extra word you find beyond **X**." I reminded participants of their scores in the first three rounds, and then showed them the actual set of letters that they would be seeing in the competition for 15 seconds prior to asking them to predict how many words they would find. I further incentivized participants' prediction accuracy by informing them: "To encourage you to predict accurately, if your prediction is within 3 words of your final score, we will give you an extra $0.03 bonus" (I chose a small sum so that participants would not stop finding words in order to win the bonus, but would at least be minimally incentivized to make accurate predictions). After seeing the competition set of letters and the score to beat, participants entered their prediction for the number of words they would find in the competition and then reported how likely it was that they were going to win the competition (1 = *not at all likely*, 10 = *very likely*), as well as whether they thought that they were going to win the competition (*Yes*, *No*).

Control variables. After the competition, I asked participants, "Overall, how did you feel about your performance in the competition?" (1 = *not at all good*, 10 = *excellent*) and "How much effort did you put into the competition?" (1 = *no effort*, 10 = *a lot of effort*), after which I revealed participants' results to them. Finally, participants completed the same measures

[9] The final method and results reported in this paper involve some deviations from this pre-registration (e.g., labeling the experimental condition "decreasing-difficulty" instead of "positive-momentum"); the deviations are described in Appendix 1.

described in Experiment 1d: task familiarity, demographics, manipulation checks, how fun the game was, and mood.

Results

Analysis plan. As in Experiments 1a–d, I removed participants who did not score within one standard deviation below the mean on their average practice round or competition score. Including these participants in analyses does not significantly change any of my results, as described in Appendix 1.

Practice round scores. Actual practice round scores were not meaningfully different between the *increasing*- (M =11.24, 95% CI [10.20, 12.29]) and *decreasing-difficulty* (M = 10.91, 95% CI [10.02, 11.80]) conditions, $t(258) = 0.48$, $p = .632$, $d = 0.06$, suggesting that my manipulation did not affect actual skill levels.

Awareness of changing performance. Indicating that participants were paying attention, they reported improving more on the practice rounds in the *decreasing-difficulty* condition (M = 1.37, 95% CI [1.17, 1.57]) than in the *increasing-difficulty* condition (M = -0.65, 95% CI [-0.89, -0.41]), $t(258) = 12.79$, $p < .001$, $d = 1.59$.

Perceived momentum. Supporting my first prediction, task-difficulty condition significantly affected momentum perceptions ($M_{\text{decreasing}} = 20.73$, 95% CI [17.33, 24.12]; $M_{\text{increasing}} = 0.24$, 95% CI [-4.25, 4.74]), $t(258) = 7.22$, $p < .001$, $d = 0.90$. Self-efficacy ratings followed a similar pattern (see Appendix 1).

Expected performance. Supporting my second prediction and replicating the findings of Experiments 1a–d, task-difficulty condition significantly influenced participants' performance expectations, such that participants in the *decreasing-difficulty* condition were more likely to believe that they were going to win the competition (M = 5.79, 95% CI [5.39, 6.19]) than were participants in the *increasing-difficulty* condition (M = 4.73, 95% CI [4.32, 5.14]), $t(258) = 3.67$, $p < .001$, $d = 0.46$. Likewise, participants in the *decreasing-difficulty* condition (M = 0.58, 95% CI [0.49, 0.66]) were more likely to believe that they would be the ultimate winner of the competition than were participants in the *increasing-difficulty* competition (M = 0.43, 95% CI [0.34, 0.52]), $t(258) = 2.37$, $p = .018$, $d = 0.30$.

Performance predictions. I next examined participants' predicted performance scores. Despite reporting that they were going to be more likely to win, participants in the *decreasing-difficulty* condition did not predict a higher competition score (M = 12.08, 95% CI [10.87, 13.28]) than did participants in the *increasing-difficulty* condition (M = 11.46, 95% CI [10.25, 12.67]), $t(258) = 0.71$, $p = .478$, $d = 0.09$.[10]

I further tested a metric more directly related to the criterion I set for winning the competition: the extent to which participants predicted that they were going to beat their target score and hence "win" the competition (i.e., the difference between participants' predicted score and average score). On this metric there was an effect of task-difficulty condition ($M_{\text{decreasing}}$ = 1.17, 95% CI [0.51, 1.83]; $M_{\text{increasing}} = 0.22$, 95% CI [-0.28, 0.72]), $t(258) = 2.26$, $p = .025$, d = 0.28, indicating that participants in the *decreasing-difficulty* condition predicted a larger margin of success than did participants in the *increasing-difficulty* condition.

Actual performance. In reality, participants were similarly likely to win in both conditions ($M_{\text{decreasing}} = 0.54$, 95% CI [0.45, 0.62]; $M_{\text{increasing}} = 0.60$, 95% CI [0.52, 0.69]), $t(258)$ = 1.03, $p = .302$, $d = 0.13$. Furthermore, performance levels did not vary by condition, ($M_{\text{decreasing}}$

[10] When controlling for average score in an ANOVA, participants in the *decreasing-difficulty* condition did predict a higher competition score than those in the *increasing-difficulty* condition, $F(1, 257) = 5.54$, $p = .019$, $\eta^2 = 0.01$. However, this was not a pre-registered analysis.

= 12.76, 95% CI [11.38, 14.14]; $M_{increasing}$ = 13.05, 95% CI [11.78, 14.31]), $t(258) = 0.31$, p = .760, d = 0.04, even when controlling for average practice score. The actual number of words that participants found above their target score (i.e., the difference between their actual competition score and average practice round score) also did not differ by condition ($M_{decreasing}$ = 1.85, 95% CI [0.70, 2.99]; $M_{increasing}$ = 1.80, 95% CI [1.19, 2.42]), $t(258) = 0.07$, p = .947, d = 0.01, even when only participants who won the competition are included in the analysis ($M_{decreasing}$ = 5.14, 95% CI [3.37, 6.91]; $M_{increasing}$ = 3.78, 95% CI [3.10, 4.46]), $t(146) = 1.48$, p = .142, d = 0.24.

Momentum miscalibration? To more fully understand and quantify discrepancies between participants' predicted and actual performance levels, I computed values corresponding to the four categories employed in signal detection theory: 1) *hit* (expect to win and win), 2) *miss* (expect to win but lose), 3) *correct rejection* (expect to lose and lose), and 4) *false alarm* (expect to lose but win). As depicted in Table 3, participants in the *decreasing-difficulty* condition (compared to participants in the *increasing-difficulty* condition) were more likely to experience *misses* ($M_{decreasing}$ = 0.24, 95% CI [0.17, 0.32], $M_{increasing}$ = 0.14, 95% CI [0.08, 0.20]), $t(258)$ = 2.09, p = .037, d = 0.26, whereas participants in the *increasing-difficulty* condition (compared to participants in the *decreasing-difficulty* condition) were more likely to experience *false alarms* ($M_{decreasing}$ = 0.20, 95% CI [0.13, 0.27], $M_{increasing}$ = 0.31, CI [0.23, 0.39]), $t(258) = -2.00$, p = .047, d = -0.25. In other words, 10.18% of participants were more likely to incorrectly predict that they would *win* in the *decreasing-difficulty* condition, whereas 10.80% were more likely to incorrectly predict that they would *lose* in the *increasing-difficulty* condition. The number of participants who correctly predicted that they would win (*hits*; $M_{decreasing}$ = 0.33, 95% CI [0.25, 0.41], $M_{increasing}$ = 0.29, 95% CI [0.21, 0.37]), $t(258) = 0.77$, p = .443, d = 0.10, or lose (*correct rejections*; $M_{decreasing}$ = 0.22, 95% CI [0.15, 0.29], $M_{increasing}$ = 0.26, 95% CI [0.18, 0.33]), $t(258)$ = -0.72, p = .473, d = -0.09, did not differ by condition.

Table 3. *Number and percentage of participants in each experimental condition who correctly or incorrectly anticipated a win or loss (N=260) in Experiment 2.*

	Decreasing Difficulty		Increasing Difficulty				
	Number of Participants in Condition	Percentage in Condition	Number of Participants in Condition	Percentage in Condition	*t*	*p*	*d*
Expect to win but lose (Miss)	32	24.24%	18	14.06%	2.09	.037	0.26
Expect to lose but win (False Alarm)	27	20.45%	40	31.25%	-2.00	.047	-0.25
Expect to win and win (Hit)	44	33.33%	37	28.90%	0.77	.443	0.10
Expect to lose and lose (Correct Rejection)	29	21.97%	33	25.78%	-0.72	.473	-0.09
Total	132	100%	128	100%			

To further test for miscalibration, I examined a pre-registered metric: the difference between participants' predicted competition scores and their actual competition scores. Only weakly consistent with my prediction, the difference was directionally but non-significantly higher in the *decreasing-difficulty* condition than in the *increasing-difficulty* condition ($M_{\text{decreasing}}$ = -0.68, 95% CI [-1.58, 0.22]; $M_{\text{increasing}}$ = -1.59, 95% CI [-2.19, -0.98]), $t(258) = 1.65$, $p = .100$, d = 0.21. Indeed, participants were unexpectedly *underconfident* in both conditions, believing that they would find fewer words than they actually did. An alternative analysis intended to control for competition score revealed similar results: In a one-way ANOVA of condition on prediction controlling for competition score, another non-significant but directionally consistent effect of condition emerged, $F(1, 257) = 2.79$, $p = .096$, $\eta^2 < 0.01$, suggesting that predictions were slightly more highly correlated with actual scores in the *decreasing-difficulty* condition than in the *increasing-difficulty* condition.

Control variables. I also tested whether task-difficulty condition affected participants' moods or beliefs about whether the task was fun. There were no differences in mood across condition ($M_{\text{decreasing}}$ = 4.62, 95% CI [4.36, 4.89]; $M_{\text{increasing}}$ = 4.59, 95% CI [4.33, 4.84]), $t(258)$ = 0.19, $p = .850$, $d = 0.02$, but participants in the *decreasing-difficulty* condition did believe that the task was more fun ($M_{\text{decreasing}}$ = 4.14, 95% CI [3.89, 4.40]; $M_{\text{increasing}}$ = 3.75, 95% CI [3.46, 4.04]), $t(258) = 2.03$, $p = .044$, $d = 0.25$. However, the aforementioned results on expectations, accuracy, and performance levels do not meaningfully change when I control for perceived fun, nor do they change when I control for participants' average practice score, task familiarity, education, age, or gender (see Appendix 1).

Discussion

Experiment 2 demonstrated with additional precision how momentum perceptions give rise to miscalibrated performance expectations and performance levels. I replicated the effect from Experiments 1a–d that momentum perceptions raise performance expectations without boosting performance levels. Moving beyond this, I also showed that performers who complete practice tasks in decreasing-difficulty order, and thereby experience gaining momentum, are also more likely to incorrectly predict that they will win. Conversely, when completing tasks in increasing-difficulty order, and thereby experiencing losing momentum, performers are more likely to incorrectly predict that they will lose. To examine whether (inflated) performance expectations have behavioral consequences, the next experiment measures performers' betting behavior.

Experiment 3: Betting on Momentum

Experiment 3 examines whether completing tasks in decreasing-difficulty order (creating the experience of gaining psychological momentum) as opposed to increasing-difficulty order (creating the experience of losing psychological momentum) will embolden people to bet on themselves in a competition, even if those who are in the former condition are objectively no more likely to win than those in the latter. Furthermore, Experiment 3 considers two possible moderators of the effect of psychological momentum on performance expectations: performer status and mindset. First, I examine how underdog (versus favorite) status influences performance expectations, as prior research suggests that observers sometimes have higher performance expectations for underdogs in a variety of contexts (e.g., Vandello, Goldschmied, & Richards, 2007). Second, I test whether performers who believe that their abilities are malleable will experience a more pronounced effect of psychological momentum on performance expectations than performers who believe that their abilities are fixed.

Method

Participants. Because I did not know exactly what effect size to expect with regard to betting behavior, I decided to collect 100 participants for each of four experimental conditions in the hope that a sample size of this magnitude would yield adequate statistical power to detect a small to medium-sized effect. In all, I recruited 409 participants (224 female, M_{age} = 38.28 years, 95% CI [37.07, 39.50]) through Amazon Mechanical Turk who completed a survey in exchange for $0.60 with the opportunity to earn a $0.40 bonus.

Procedure. The study employed a 2 (task order: increasing-difficulty or decreasing-difficulty) × 2 (status: favorite or underdog) between-participants design. The procedure and survey followed Experiment 1d except for the following changes described below.

Growth mindset. Prior to completing the practice rounds, participants completed a growth mindset scale (eight items, modified to pertain to performance; De Castella & Byrne, 2015).

Favorite vs. underdog. After completing the practice rounds, I randomly assigned participants into a status condition (*favorite* or *underdog*) and told them: "You have the opportunity to play one more round and earn a bonus if you would like to. In this round you will compete against a set score." Participants in the *underdog* condition learned that they needed to beat their average score *plus one* (thereby making it harder for them to win), whereas participants in the *favorite* condition learned that they needed to beat their average score *minus one* (thereby making it easier for them to win).

Betting on performance. To explain the choice to enter the competition, participants read:

> "If you WIN the competition you will earn **an additional $0.40**. If you LOSE the competition, you will earn **$0**. You do not get ANY bonus money if you lose the competition. If you choose **not** to compete, you will not play the next round, and instead just answer a few questions and finish this experiment. If you do not choose to compete, you will earn **an additional $0.20** which cannot be taken away. To summarize, your choices are: 1. Choose to compete: win the competition = $0.40 extra, lose the competition = $0.00 2. Choose not to compete = $0.20 extra."

I then reminded participants of their practice scores in each round and the score they needed to beat to win the competition, after which they reported how likely they believed it was going to be that they would win the competition (1 = *not at all likely*, 10 = *very likely*). For my key betting measure, participants selected whether or not they wanted to enter the competition (*Yes*, *No*).

Control variables. Following the competition, participants completed the same measures described in Experiment 1d: task familiarity, demographics, manipulation checks, whether the game was fun, and their mood.

Awareness of changing task difficulty. Finally, in addition to examining participants' awareness of how their own performance was changing, this experiment also measured participants' awareness that the task difficulty was changing (i.e., an external attribution for their momentum). To examine this, I asked: "Did you notice that the word sets were getting easier or harder during the first three rounds?" (1 = *Yes, they were getting easier*, 2 = *Yes, they were getting harder*, 3 = *No, I didn't notice a change*).

Results

Analysis plan. As in Experiments 1a–d and 2, I removed participants who did not score within one standard deviation below the mean on their average practice round score or competition score. Including these participants in analyses does not significantly change any of my results, as described in Appendix 1.

Practice round scores. Average practice scores were higher in the *decreasing-difficulty* condition (M = 12.26, 95% CI [11.52, 13.01]) than in the *increasing-difficulty* condition (M = 11.10, 95% CI [10.40, 11.79]), $t(342) = 2.27$, $p = .024$, $d = 0.25$. This difference was driven solely by a difference in letters found on their easy set ($M_{\text{decreasing}}$ = 15.88, 95% CI [14.83, 16.93]; $M_{\text{increasing}}$ = 12.13, 95% CI [11.29, 12.98]), $t(342) = 5.53$, $p < .001$, $d = 0.60$. The averages of the medium and hard sets of letters were the same across condition, ts < 1.40, ps > 0.162, ds < 0.16. Because the participants in the *decreasing-difficulty* condition saw the easy set of letters last, I believe this difference is a practice effect and not a skill effect.

Awareness of changing performance. As in prior experiments, participants reported their performance was improving more on the practice rounds in the *decreasing-difficulty* condition (M = 1.40, 95% CI [1.23, 1.57]) than in the *increasing-difficulty* condition (M = -0.66, 95% CI [-0.89, -0.41]), $t(342) = 13.97$, $p < .001$, $d = 1.51$.

Awareness of changing task difficulty. Those in the *increasing-difficulty* condition recognized that the tasks were getting harder (M = 1.17, 95% CI [1.09, 1.25]), whereas those in the *decreasing-difficulty* condition recognized that the tasks were getting easier (M = 1.85, 95% CI [1.79, 1.91]), ps < .001, and the two conditions reported significantly different assessments of task-difficulty, $t(221) = 13.78$, $p < .001$, $d = 1.85$. Thus, participants were aware that the task difficulty was changing.

Perceived momentum. Replicating the results of the prior experiments and supporting my first prediction, perceived momentum measured prior to the competition task varied by task-difficulty condition ($M_{\text{decreasing}}$ = 64.52, 95% CI [60.85, 68.19]; $M_{\text{increasing}}$ = 45.21, 95% CI [41.02, 49.40]), $t(342) = 6.80$, $p < .001$, $d = 0.74$. I did not analyze momentum scores by status because participants provided their momentum perceptions before I assigned them to different statuses.

Expected performance. Replicating the results of the prior experiments and supporting my second prediction, a 2 × 2 ANOVA conducted on performance expectations indicated that participants in the *decreasing-difficulty* condition were more likely to believe that they were going to win (M = 6.13, 95% CI [5.80, 6.46]) than were participants in the *increasing-difficulty* condition (M = 5.38, 95% CI [5.04, 5.73]), $F(1, 340) = 12.29$, $p < .001$, $\eta^2 = 0.03$. Unsurprisingly, *favorites* also believed that they were going to be more likely to win (M = 6.27, 95% CI [5.92, 6.62]) than did *underdogs* (M = 5.25, 95% CI [4.93, 5.57]), $F(1, 340) = 18.43$, $p < .001$, $\eta^2 = 0.06$, but there was no interaction, $F(1, 340) = 0.39$, $p = .531$, $\eta^2 < 0.01$.

Momentum mediates performance expectations. As I found in Experiments 1a–d and Experiment 2, perceived momentum mediated the effect of task-difficulty condition on performance expectations in a 10,000 bootstrapping mediation model, 95% bias-corrected CI [0.33, 0.69]. In a separate model, momentum perceptions did not mediate the effect of status condition on performance expectations, 95% bias-corrected CI [-0.21, 0.07].

Betting on performance. Testing my hypothesis that task order would influence betting behavior, I found that participants were indeed more likely to choose to enter the competition in the *decreasing-difficulty* condition (M = 0.78, 95% CI [0.72, 0.85]) than they were in the *increasing-difficulty* condition (M = 0.67, 95% CI [0.60, 0.74]), $F(1, 340) = 6.52$, $p = .011$, $\eta^2 = 0.02$. Participants were also more likely to choose to enter the competition in the *favorite* condition (M = 0.80, 95% CI [0.74, 0.86]) than in the *underdog* condition (M = 0.66, 95% CI

[0.59, 0.73]), $F(1, 340) = 8.58$, $p = .004$, $\eta^2 = 0.03$. The interaction was non-significant, $F(1, 340) = 0.27$, $p = .607$, $\eta^2 < 0.01$.

Actual performance. I next examined competition performance, noting that because participants self-selected into the competition (78% of *decreasing-difficulty* participants and 67% of *increasing-difficulty* participants chose to compete), random assignment to condition no longer holds in subsequent analyses. Despite clear differences in performance expectations, as well as an expressed willingness to *bet* on those expectations, neither task-difficulty condition nor status condition affected performance levels ($M_{\text{inc-fav}} = 11.13$, 95% CI [9.46, 12.80]; $M_{\text{inc-fav}} = 10.05$, 95% CI [7.63, 12.46]; $M_{\text{dec-fav}} = 13.50$, 95% CI [11.48, 15.52]; $M_{\text{dec-under}} = 11.39$, 95% CI [9.43, 13.35]) after controlling for average practice scores, $F(1, 339) = 0.13$, $p = .714$, $\eta^2 < 0.01$ and $F(1, 339) = 2.69$, $p = .102$, $\eta^2 < 0.01$, respectively, and there was no interaction, $F(1, 339) = 1.16$, $p = .282$, $\eta^2 < 0.01$. However, *favorites* actually "won" the competition more often ($M = 0.87$, 95% CI [0.81, 0.92]) than did *underdogs* ($M = 0.75$, 95% CI [0.67, 0.82]), presumably because they were competing against a lower score, $F(1, 246) = 5.33$, $p = .022$, $\eta^2 = 0.02$.

Growth mindset and performance expectations. To examine the relationship between growth mindset and performance expectations, I ran a regression analysis with expectations as the dependent variable that included the following predictors: task-difficulty condition (0 or 1), status condition (0 or 1), growth mindset score, all interactions between these variables, and average practice score. The analysis revealed a positive relationship between performance expectations and growth mindset, $\beta = 0.49$, 95% CI [0.10, 0.89], $p = .015$, indicating that a tendency to believe in the malleability of one's performance abilities was associated with higher performance expectations, and between performance expectations and practice score, $\beta = 0.05$, 95% CI [0.00, 0.10], $p = .045$, but no other main effects or interactions. Removing the interaction variables in a second analysis revealed relationships between performance expectations and task-difficulty condition ($\beta = 0.94$, 95% CI [0.49, 1.39], $p < .001$), status condition ($\beta = 1.18$, 95% CI [0.74, 1.63], $p < .001$), growth mindset ($\beta = 0.58$, 95% CI [0.38, 0.78], $p < .001$), and, marginally, practice score ($\beta = 0.04$, 95% CI [-0.01, 0.09], $p = .082$).

Growth mindset and actual performance. To examine the relationship between growth mindset and performance levels, I re-ran the two regression analyses described above, substituting performance levels for performance expectations as the dependent variable. The first analysis included the following predictors: task-difficulty condition (0 or 1), status condition (0 or 1), growth mindset score, all interactions between these variables, and average practice score. In this analysis, none of the variables predicted performance levels except for average practice score ($\beta = 1.31$, 95% CI [1.16, 1.47], $p < .001$). Removing the interaction variables in a second analysis revealed a relationship between growth mindset and performance, indicating that the more malleability participants perceived in their performance abilities, the better they performed in the competition ($\beta = 1.12$, 95% CI [0.46, 1.78], $p < .001$), as well as a (marginal) effect of status on performance ($\beta = 1.46$, 95% CI [0.02, 2.93], $p = .053$), indicating that *favorites* ($M = 12.17$, 95% CI [10.89, 13.46]) scored higher in the competition than did *underdogs* ($M = 10.75$, 95% CI [9.22, 12.28]).

Discussion

Overall, Experiment 3 replicates the findings obtained in the previous experiments, and also shows that performers are more likely to bet on themselves when they believe they are gaining (versus losing) momentum, even though their objective performance does not differ between conditions. Thus, inaccurate financial gambles may be one potential consequence of momentum miscalibration.

I observed no interactions between the experience of gaining or losing momentum and performers' growth mindset or status (i.e., favorite or underdog) on either expected or actual performance. This suggests that the effect of momentum on miscalibrating expectations is not meaningfully influenced by performers' mindset or status. However, growth mindset and status both directly influenced expectations and real performance: participants who believed their performance was more malleable, and those who were actual favorites, had higher expectations and performed better.

Finally, this experiment provides more insight into people's beliefs about the changing difficulty-order of the practice tasks they completed. As in prior experiments, performers reported awareness that they were getting better or worse over time, consistent with the difficulty ordering of the tasks. But in this experiment, performers further reported that they knew that the task difficulty level was changing. This suggests that although performers have at least some knowledge about the source of their experienced momentum, the feeling of momentum can still elicit miscalibration.

Discussion (Chapter 1)

Momentum is a significant psychological force, and these findings suggest that it exerts a stronger effect on people's performance expectations than it does on their performance outcomes. Psychological momentum can therefore be *miscalibrating*, misaligning beliefs with reality. Six experiments ($N = 2{,}533$) revealed the remarkably consistent result that experiencing momentum in a competitive domain can enhance performance *expectations* without actually enhancing performance *outcomes*. Moreover, this effect of perceived momentum on expectations, as well as its corresponding null effect on performance, was robust to players' skill levels, task familiarity, self-efficacy levels, growth mindset, favorite versus underdog status, and various demographic variables.

My research provides a quantifiable and robust measure of the influence of momentum perceptions on calibration. For instance, in Experiment 2 I found that performers who completed tasks in decreasing-difficulty order were 10.18% more likely than performers who completed tasks in increasing-difficulty order to predict that they would *win* a competition that they eventually *lost*. Momentum miscalibration can have meaningful consequences: in Experiment 3, performers' miscalibrated expectations led them to financially bet on their competition performance more than they should have. In other words, players sacrificed their own earnings to bet on their performance when they felt they were gaining momentum.

Theoretical Contributions

I believe that the present work makes at least four theoretical contributions. First, it marks an initial effort to quantify the extent to which momentum perceptions give rise to miscalibrated performance expectations. Whereas prior research has provided hints about the nature of discrepancies between expectations and performance outcomes under conditions of perceived momentum (Gilovich et al., 1985; Perreault et al., 1998; Shaw et al., 1992), none have examined or measured the magnitude of miscalibration under varying levels of (manipulated) momentum.

Second, I contribute to a growing body of research examining the consequences of momentum and momentum-like experiences on beliefs about the future. Most relevant is the prior finding that telling people to be optimistic or confident can enhance expectations about their future performance (Tenney et al., 2015). Similar results have been observed in perceptions of status within groups and probability assessments. In one set of experiments, individuals who ascended through a hierarchy to reach their current rank were rated more positively and given more esteem than those who fell to that same rank (Pettit, Sivanathan, Gladstone, & Marr, 2013), suggesting that people expect changes in social hierarchy to continue linearly. In another set of experiments, when the probability of an event increased, the event felt less remote and more likely to occur than when the probability decreased (even when the final probability was exactly the same; Maglio & Polman, 2016). Extending this research, I find that not just observing but also *experiencing* momentum changes beliefs about the future.

Third, I provide an additional and rigorous test of whether momentum can influence performance outcomes in a domain that requires both skill and effort. Although a number of theories propose that momentum should influence performance outcomes under the very conditions I tested (Adler, 1981; Briki et al., 2014; Iso-Ahola & Dotson, 2016), I find no evidence for a causal effect of momentum on performance. However, as I note in the future directions section below, this is an important area that future research should explore.

Finally, I demonstrate that momentum perceptions not only inflate people's performance expectations, but also lead people to *act* on their inflated expectations. Whereas prior research has examined correlations between momentum and betting behavior for other actors (Arkes & Martinez, 2011; Jegadeesh & Titman, 1993), I provide an initial demonstration that psychological momentum can causally increase betting on oneself.

Limitations and Future Directions

This research contains limitations that future work could explore. One important question is whether, and under what conditions, momentum actually influences performance. In the present work I found no clear evidence that momentum causally affects performance, and the null effect that I observed was adequately statistically-powered to detect an effect if it had existed. Of course, as with any null effect, my finding is not conclusive, although it is consistent with a number of other reported null findings (Gilovich et al., 1985; Mack & Stephens, 2000; Shaw et al., 1992; Silva, Cornelius, & Finch, 1992). On the other hand, my continuous measure of perceived momentum did positively predict performance levels in five out of six experiments while controlling for skill level and other relevant variables, indicating that perhaps momentum does play a role that I did not fully capture in my experimental paradigm.

It is also possible that I did not examine the performance domains or conditions under which momentum might actually influence performance. For instance, Taylor and Demick (1994) suggest that physiological arousal and positive affect are necessary for momentum to impact performance, and Vallerand et al. (1988) claim that some degree of control perceptions are necessary for the momentum-performance effect to appear. One recent paper (Shen & Hsee, 2017) demonstrated that viewing incidentally increasing numbers in increasing (vs. constant or decreasing) velocity made people perform more quickly when identifying words in a computer game and even step more quickly on a stepping machine, an effect that may be partly driven by the perception of momentum (although this perception was not measured in the reported studies). Thus, future research might examine the momentum-performance link by creating conditions that incite physiological arousal and positive affect, and employ tasks that elicit control perceptions in order to more thoroughly test these theories.

Second, another direction for future research is to examine how experiencing momentum influences motivation. For instance, might ordering daily tasks from difficult to easy lead people to believe that they are gaining momentum throughout the day? Or, might people benefit from setting more challenging goals at first (and easier goals later) in order to establish and build momentum? This possibility would diverge somewhat from predictions made by goal-setting theories, which posit that setting more challenging goals should lead people to perform better (Locke & Latham, 2002). Thus, researchers might benefit from integrating momentum and goal-setting theories in order to better understand how people experience progress versus setbacks when they are striving to attain their goals (e.g., Koo & Fishbach, 2014).

Third, there may be many more unexplored consequences of momentum miscalibration. Experiencing or perceiving psychological momentum could influence decisions across a variety of domains. For example, momentum could lead individuals to purchase items that they might not necessarily need (e.g., impulse buying; Dhar, Huber, & Khan, 2007), fall prey to emotional contagion (e.g., "auction fever"; Ku, Malhotra, & Murnighan, 2005), and make risky choices in domains such as stock investing and gambling where control perceptions are mostly illusory (Guenther & Kokotajlo, 2017). But it may also be possible to counter such tendencies, for instance by reminding individuals of base-rate frequencies and prior probabilities, or perhaps by guiding people to attribute momentum externally (e.g., to a changing situation) instead of internally (e.g., to their own skill levels).

I also encourage researchers to further examine the psychological underpinnings of momentum miscalibration. Is momentum miscalibration a cognitive bias whereby performers perceive exaggerated causal connections between the trajectories of their recent performance outcomes and their future performance outcomes (cf. Tormola, Jia, & Norton, 2012), or could it be motivated by control or sense-making needs? Why has a tendency for people to infer over-inflated assessments of their capabilities from transient, phenomenological experiences emerged? Exploring the psychological reasons for why momentum miscalibration exists could be a fruitful avenue for future research.

Finally, I am curious whether there are boundary conditions to the link between momentum and performance expectations. In all of the experiments in this chapter, momentum correlated very highly with self-efficacy. I did not explore situations in which self-efficacy does not vary with momentum, but perhaps these two constructs are so closely linked that perceived momentum only predicts performance expectations when momentum signals some sort of efficacy. I plan to explore this question more thoroughly in Chapter 2.

CHAPTER 2:

Efficacious Momentum:
Momentum Leads to Expectations to Win, but Only for Skilled Actors

Abstract (Chapter 2)

The "hot hand" phenomenon predicts that a string of successes will lead to continued success, but the gambler's fallacy predicts that a string of successes will lead to a reversal of that success. This research attempts to reconcile those opposing predictions by exploring situations in which an agent is gaining or losing momentum and collecting observers' beliefs about that agent's subsequent performance. I predict that people believe that positive trajectories will continue when the changing momentum signals the target's underlying efficacy, such as in performance domains, but will believe that positive trajectories will reverse when momentum seems to be driven by luck or when efficacy is not present. In Experiment 1 ($N = 425$), knowledge of recent positive or negative movement through rankings changed performance expectations for closely ranked competitors, mediated by ratings of efficacy. In Experiment 2 ($N = 407$), participants believed so strongly in momentum's ability to improve performance that they were even willing to bet money on an objectively lower-ranked performer when told that the lower-ranked performer had been gaining momentum and the higher-ranked performer had been losing momentum. In Experiment 3 ($N = 307$), I provided observers with a non-efficacious attribution for the actor's changing trajectory, which turned off the effect of momentum on performance predictions. Finally, in Experiment 4 ($N = 208$), I manipulated whether the change in trajectory seemed to be driven by skill (i.e., free throws) or luck (i.e., roulette), and found that the momentum attribution interacts with predictions in that participants expect performers in high-efficacy domains to continue on the same trajectory and those in low-efficacy domains to reverse trajectory. Together, these experiments provide insight on how people extrapolate trajectories relative to gaining or losing momentum as a function of whether the trajectory seems to be due to luck or skill.

Imagine watching a basketball player sink three shots in a row; will his next shot be a hit or a miss? The player's sequence of successful shots creates a perception among observers that the player is gaining momentum, the noticed change in upward or downward trajectory in pursuit of a goal. Humans are predisposed to seek patterns in every sequence, even sequences of random events (Shermer, 2008). For example, just consider the beliefs that people form around a coin flip that continually lands on "heads" instead of "tails." Even if they know that the outcome of each flip is independent from the last, they form beliefs about the flipper, the coin, or the environment to help explain a seemingly unlikely sequence of events (Gilovich, 1991). People then use those patterns as information to predict what will happen next. This is especially relevant in competition settings—contests in which one person or team wins and the other loses.

But predictions about momentum are hardly uniform. Consider the contrasting beliefs of the "hot hand" phenomenon (e.g., Gilovich, Vallone, & Tversky, 1985), in which people expect a series of successes to result in future success, versus the gambler's fallacy and regression to the mean (Tversky & Kahneman, 1971), in which people expect a series of successes to result in future failures. I propose a way to integrate these seemingly conflicting literatures by suggesting that a critical difference between these two well-documented phenomena are the inferences that the observer forms about the player's efficacy. When a player's efficacy—that is, their beliefs in their capabilities (Bandura, 1977), which I operationalize as a combination of skill, confidence, and trust in abilities—can conceivably affect their likelihood of another success, momentum seems to breed expectations of future success (e.g., the player has a "hot hand"; Attali, 2013; Croson & Sudali, 2005; Gilovich et al., 1985; Markman & Guenther, 2007). But conversely when the success involves a more externalized attribution, such as luck, God, or statistics, momentum may seem quite unlikely to increase future success (Ayton & Fischer, 2004; Burns & Corpus, 2004; Tversky & Kahneman, 1971). Instead, people may even predict future performance will decline because the player's "luck has run out" or because they seem otherwise due to regress back toward a more reasonable average.

Defining Momentum

In Chapter 1, I manipulated feelings of momentum in performers and I defined psychological momentum as the felt progress of moving toward or away from a goal preceded by an experienced change in upward or downward trajectory in pursuit of that goal. In this chapter, because I focus on observers' perceived momentum, I define psychological momentum as the observed movement toward or away from a goal preceded by an observed change in upward or downward trajectory in pursuit of that goal.

In this chapter I manipulate momentum in two ways. In the first three experiments, I manipulate perceived momentum by varying the amount of information observers have about an actors' movement through a set of rankings. Some participants learn that the targets had recently moved through the rankings and others are only told the targets' current ranks. In the fourth experiment, I manipulate momentum by telling participants about the actor's current streak of either successes or losses. In both of these manipulations participants use trajectory information to infer momentum.

Momentum and Expectations About the Future

When a person observes positive momentum in a performer, it can affect their beliefs about the target's future trajectory in two possible ways. First, they might believe the positive trajectory will continue or even increase, such that the performer's future trajectory looks similar to the performer's past trajectory (e.g., hot hand). But second, alternatively, they might believe the trajectory will reverse such that the future trajectory is the opposite of the past trajectory (e.g., regression to the mean, gambler's fallacy). When does a sequence of successes lead to each of those predictions? I propose that in order to predict a streak will continue on its current trajectory, the initial streak must be driven by some perception of efficacy.

To my knowledge, at least two research papers have made consistent assertions. Croson and Sundali (2005) suggest "individuals who believe in the hot hand believe not that a particular outcome is hot (e.g. that the roulette wheel that has come up red in the past is likely to come up red again), but that a particular *person* is hot" (p. 198). Similarly, Ayton and Fischer (2004) "propose that sequences of outcomes reflecting human performance yield anticipations of positive recency, whereas outcomes due to inanimate chance mechanisms yield anticipations of negative recency" (p. 1374). I take these explanations one step further and argue that the reason the hot hand phenomenon only applies to people is that it requires an inference about an actor's efficacy. This further suggests that removing efficacy should remove the effect of momentum on performance expectations.

Predicting failure after successes. The gambler's fallacy (Tversky & Kahneman, 1971) and the concept of regression towards the mean are phenomena in which people expect a reversal in trajectory. The gambler's fallacy holds that if something happens more frequently than "normal" (which is subjectively determined by the perceiver) during some period, it will happen less frequently in the future. At horse races, for instance, people are less likely to bet on the favorite if the favorite has won the last two races, even though they are completely different animals (Metzger, 1984). In fact, I suggest it is actually *because* they are completely different animals. If the previous wins signaled efficacy[11] for the horse and jockey that were currently racing, people would probably bet on that horse more often, not less. Instead, people are attributing the streak of successes to the label of favorite. Since a label cannot signal efficacy, people predict that the streak will reverse. People are also less likely to play lottery numbers that have won recently (Clotfelter & Cook, 1991, 1993; Terrell, 1994) and more likely to play lottery numbers that have not been picked for a long time (Oskarsson, Van Boven, McClelland, & Hastie, 2009) despite the fact that each lottery selection is an independent event. The implication is that those numbers have experienced some success (or failure), but with no efficacy present, people expect those numbers to experience subsequent failure (or success).

Because roulette is a game of randomness, it provides many examples of the gambler's fallacy in action. In one experiment, Ayton and Fischer (2004) spun a simplified roulette wheel 200 times and after each spin asked participants to predict which color would appear next. In line with the gambler's fallacy, they found that the longer a run of a particular color was, the less likely participants were to choose that color for the next spin. In a different experiment, only 12% of people predict a streak of four reds will continue when predicting the fifth spin (Burns & Corpus, 2004). The other 88% think the fifth spin will be black even though each spin is actually independent from the prior spin (i.e., 50% likelihood of black). Researchers also found evidence of the gambler's fallacy when analyzing 18 hours of data from a casino's roulette table. They discovered that after streaks of five or more spins that landed on red/blue, even/odd, or low/high numbers, gamblers were significantly more likely to bet against the streak than with it (Croson & Sundali, 2005).

Predicting success after successes. The aforementioned work shows examples in which people believe a positive trajectory predicts a future negative trajectory. It is also possible that the future trajectory could continue along the same path or increase. I point to two bodies of

---

[11] Specifically, in this case, I refer to skill, since it is arguable whether horses can have confidence or trust in their abilities, though their jockeys certainly can.

literature that support this expectation: correlational data connecting momentum to positive performance expectations and experiments that manipulate momentum-like experiences.

A paradigmatic example of the belief that momentum can positively predict future performance comes from the "hot hand" effect (Gilovich et al., 1985), which I discuss thoroughly in Chapter 1. The findings in this field highlight individuals' beliefs that a trajectory will continue, suggesting that momentum will positively affect future performance beliefs. However, most of the studies make no effort to disentangle efficacy from momentum, so a salient alternative explanation of why performance beliefs seemed to change after a string of successes is the actual differences in skill of the players. A player who wins more is likely to be a genuinely more skilled player than a player who wins less.

One study did attempt to isolate momentum as a mechanism by controlling for skill. Researchers showed participants footage from a basketball game in which the trailing team was gaining on the leading team while the leading team's point total remained stagnant. Participants thought the trailing team had more momentum and would end up winning even when their point total was still lower than the leading team's (Markman & Guenther, 2007). Since participants in this study looked at the same two teams in the same time frame, the teams' actual skill remained relatively constant. However, these findings are likely still driven by a difference in perceived efficacy since confidence and trust in abilities can improve even as objective skill remains constant.

Experiments that manipulate momentum-like experiences show a similar pattern of results. For instance, in the domain of status rankings instead of performance rankings, people expect changes in social hierarchy to continue linearly, such that individuals who ascended through a hierarchy to reach their current rank were rated higher and given more esteem than those who had fallen to reach that same rank (Pettit et al., 2013). In performance domains, components of efficacy such as confidence and optimism can affect performance expectations (Moore & Healy, 2008, Tenney et al., 2015). For example, people tend to prescribe optimism because they believe it will improve performance (the optimism-performance hypothesis; Tenney et al., 2015). In one experiment, participants who had been told they performed above average (70% correct) on a task expected to do better than those who had been told they performed below average (30% correct). Both the actors and the observers expected the "optimistic" group to perform better, but the performances were no different between groups. These results point toward a causal link between momentum and positive performance expectations but do not directly manipulate momentum or investigate the effect of efficacy on the predictions.

To my knowledge, only a few experiments have directly manipulated momentum and measured the performance expectations of observers. Participants who watched a graphic of a cyclist who was manipulated to be either gaining or losing momentum reported a belief that the cyclist would continue on their current trajectory (Briki et al., 2014). In a different experiment participants were told about a tennis match in which the score was 5-5. They were shown a game pattern that suggested the players had been fairly even up to that point, mostly trading games (the no momentum condition), or that one player had won the most recent four games (the momentum condition). Participants in the momentum condition expected the player who won the last game to win the match significantly more often than did those in the no momentum condition (Vallerand et al., 1988). These experiments are valuable and seem to causally link momentum to expected performance, but like most of the early experiments in this field, they fail to disentangle momentum from skill and suffer from low sample size (e.g., Vallerand et al. had

11 participants per condition). Therefore, I believe that the distinct roles of momentum and efficacy on expected performance deserve further investigation.

Overview of Studies

The current research uses experimental manipulations of gaining and losing momentum to test the links between perceived momentum, perceived efficacy, and expected performance. First, Experiments 1 and 2 test whether the effect of momentum on performance expectations is mediated by perceptions of performers' efficacy. In Experiment 1, I test whether differing information about ranking movement influences participants' perception of players' momentum, and subsequently players' efficacy, and consequently affects observers' expectations about who will win a match. In Experiment 2, which serves as a robustness check for the initial effect, I incentivize participants to be honest about their expectations by offering a bonus for predicting the correct winner. Since Experiments 1 and 2 are both high-efficacy domains, I expect participants to predict the actors' trajectories will continue. Next, Experiments 3 and 4 test whether limiting players' efficacy removes the effect of momentum on performance expectations. In Experiment 3, I manipulate efficacy by offering an alternative explanation for players' trajectories than the actors' own actions. This allows me to investigate whether efficacy is necessary for momentum to influence performance expectations. Experiment 4 examines diverging predictions about the effect of gaining momentum on subsequent success by comparing a domain that seems to require a player's efficacy (e.g., a basketball game) to a domain that does not seem to require as much efficacy (e.g., a roulette game). I predict that perceived efficacy will mediate the effect of gaining (versus losing) momentum on predicted future success but only in the basketball game; in the roulette game, gaining momentum will reduce expectation of future success, mediated by beliefs about luck and not about efficacy. In sum, Experiments 1 and 2 will test the link between momentum and expectation to win in a domain where gaining momentum typically signals a player's efficacy and Experiments 3 and 4 will explore whether efficacy is necessary to predict continued streaks.

Experiment 1: Perceived Momentum and Performance Expectations

I test whether manipulating how much information participants have about ranking movement influences their perception of players' momentum, efficacy, and their expectation of winning. I use actual, closely-ranked tennis players who have recently risen or fallen the same amount in the rankings and ask participants to make predictions about who will win in a match between the two. For ease of labeling, I will refer to the lower-ranked player as the underdog and the higher-ranked player as the favorite. To manipulate momentum, I gave participants either no information about each player's movement (only reporting their current rankings), information that the underdog had recently risen in the rankings (gaining momentum), information that the favorite had recently fallen in the rankings (losing momentum), or information that the underdog had recently risen and the favorite had recently fallen (both gaining and losing momentum). This design allows me to test how different levels of momentum affect beliefs about a player's efficacy and subsequent performance, in this case, whether the player will win a competition.

Method

Participants. I predetermined a sample size of 100 participants in each of four experimental conditions. In total, I recruited 425 participants (201 female, 2 gender non-binary, M_{age} = 36.37 years, 95% CI [33.23, 39.51]) through Amazon Mechanical Turk who completed a survey in exchange for $0.40.

Experimental design. I randomly assigned participants to one of four conditions that manipulated how much information they received about movement in rankings: both players static, favorite falling, underdog rising, and both players moving. In the *both static* condition, I gave participants information about only the current ranking of each player just before the match, thereby providing only static information about their rankings. In the *favorite falling* condition, I told participants that the favorite had recently fallen to his current ranking, but I provided only the current ranking of the underdog with no movement information. In the *underdog rising* condition, I told participants that the underdog had recently risen to his current ranking, but gave the current ranking with no movement information for the favorite. In the *both moving* condition, I told participants that both players had recently moved through the rankings to reach their current rankings.

Procedure. I gave participants ranking information about two tennis players in accordance with the experimental design described above and asked to make predictions about the outcome of a match between them. I chose tennis players Milos Raonic (the favorite) and Gilles Simon (the underdog) because they were ranked closely to one another and they had moved significantly and symmetrically in the past year (favorite had fallen from 8th to 14th, underdog had risen from 21st to 15th). I used real players to remove any deception from the experiment. I manipulated perceived momentum by providing different information about each player's change in rankings as described in the experimental design. To see the stimuli for each condition, please see Appendix 2. After learning about the players' rankings, participants completed attention checks (see Appendix 2 for all attention checks and results in Experiments 1–3).

Expectation to win. To measure the primary dependent variable, I asked, "Who do you expect to win the match?" (0 = *Raonic*, 1 = *Simon*).

Perceived momentum. To measure perceived momentum, I asked, "Please indicate using the slider who you think has more momentum. No movement of the slider indicates equal momentum" and provided a slider scale with the favorite on one side (0 = *Raonic*) and the underdog on the other (10 = *Simon*). The slider was set by default to 5.

Perceived efficacy. To measure perceived efficacy of each player, participants completed three measures for each player entering the match: "How skilled do you think the player is?", "How much do you think the player trusts their skills?", and "How confident do you think the player is?" (1 = n*ot at all*, 7 = *very much*; α = .86 for the favorite and .83 for the underdog). I collapsed these three questions into one efficacy index per player for analysis.

Exploratory questions. At the end of the survey, after collecting the variables of interest, I asked three exploratory questions: "Who do you think Raonic expects to win the match?", "Who do you think Simon expects to win the match?", and "Who do you *want* to win the match?" (0 = *Raonic*, 1 = *Simon*).

Control variables. Finally, I asked three questions to control for expertise in tennis: "How familiar are you with professional tennis?" (*Extremely familiar*, *Very familiar*, *Moderately familiar*, *Slightly familiar*, *Not familiar at all*), "Do you follow professional tennis?" (*Always*, *Sometimes*, *Never*), and "Do you play tennis?" (*Yes, competitively*, *Yes, recreationally*, *No, but I used to*, *No, I have never played tennis*) and collected demographic information.

Results and Discussion

Expectation to win. For ease of analysis, I refer to Raonic as the favorite and Simon as the underdog. The momentum manipulation significantly affected expectation to win, $F(3, 421) = 17.55$, $p < .001$, $\eta^2 = 0.11$ (see Figure 1). As predicted, expectations that the underdog would win were lowest in the *both static* condition: participants believed that the underdog would lose ($M = 0.21$, 95% CI [0.13, 0.29]), and this condition showed lower expectations than each of the other three conditions, *t*s > 4.98, *p*s < .001, *d*s > 0.68. Conversely, in the *both moving* condition (*M* = 0.65, 95% CI [0.56, 0.74]), the expectation for the underdog to win was not only significantly higher than in the *both static* condition, but also significantly higher than the scale mid-point of 0.5, suggesting that participants believed the underdog would win even though he was currently ranked lower than the favorite, *one-sample* $t(102) = 3.19$, $p < .001$, $d = 0.63$. The *underdog rising* condition (*M* = 0.57, 95% CI [0.47, 0.66]) was no different from the *both moving* condition, $t(207) = 1.25$, $p = .213$, $d = 0.17$, but the *favorite falling* condition (*M* = 0.52, 95% CI [0.42, 0.61]) was unexpectedly different from the *both moving* condition, albeit the effect size was small, $t(211) = 1.97$, $p = .050$, $d = 0.27$. Expectations were not significantly different in the *favorite falling* and *underdog rising* conditions, $t(214) = .703$, $p = .483$, $d = 0.10$.

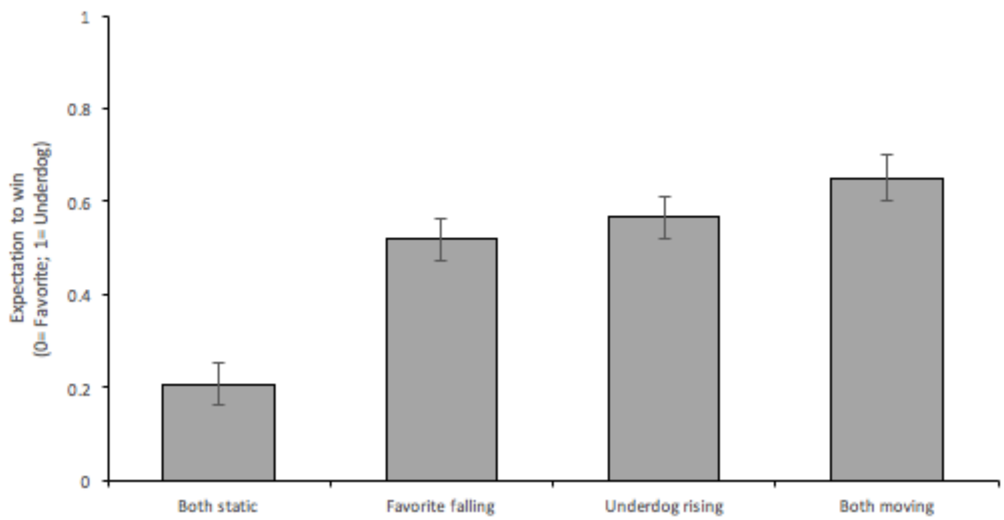


Figure 1. Expectation to win by experimental condition in Experiment 1. Error bars represent ± one standard error around the mean.

Perceived momentum. These changes in expectations were reflected in beliefs about the underdog's momentum, $F(3, 421) = 35.62$, $p < .001$, $\eta^2 = 0.20$ (see Figure 2). In particular, in the *both static* condition (*M* = 4.21, 95% CI [3.86, 4.55]), the underdog was judged to have less momentum than in any of the other conditions, *t*s > 5.75, *p*s < .001, *d*s > 0.80. As with expectation to win, the underdog seemed to have the most momentum in the *both moving* condition (*M* = 7.18, 95% CI [6.73, 7.64]), and the *underdog rising* condition (*M* = 6.93, 95% CI [6.44, 7.43]) was again no different from the *both moving* condition, $t(207) = 0.74$, $p = .462$, $d = 0.01$, perhaps because both conditions contained the same information that the underdog had recently risen in the rankings. Participants rated the underdog as having less momentum in the *favorite falling* condition (*M* = 5.92, 95% CI [5.45, 6.39]) than in the *both moving* and *underdog rising* conditions, and more than in the *both static* condition, *t*s > 2.49, *p*s < .004, *d*s > 0.35.

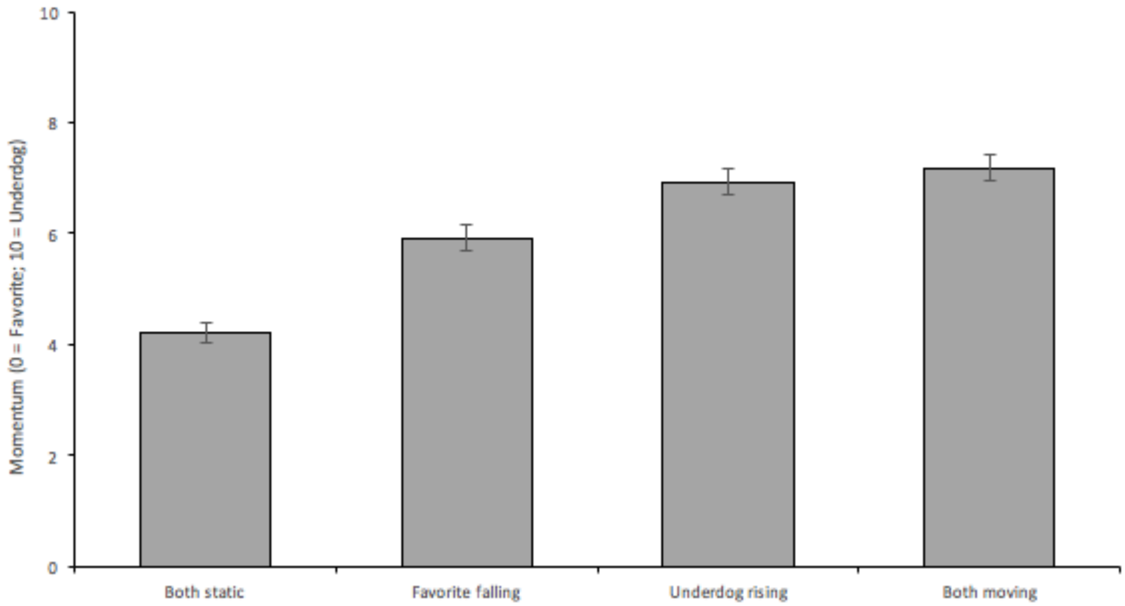


Figure 2. Perceived momentum by experimental condition in Experiment 1. Error bars represent ± one standard error around the mean.

Perceived efficacy. Furthermore, beliefs about the players' efficacy followed a similar pattern. I measured efficacy separately for each player but report it here as a difference score for ease of understanding and analysis. As with momentum and expectation to win, a higher score indicates higher efficacy for the underdog relative to the favorite. Each player's perceived efficacy was highly correlated with the other player's perceived efficacy, $r(423) = 0.49$, $p < .001$. Just as on expectations and momentum judgments, condition also had a significant effect on efficacy judgments, $F(3, 421) = 20.5$, $p < .001$, $\eta^2 = 0.13$. In the *both static* condition, the difference was significantly less than zero ($M = -0.25$, 95% CI [-0.48, -0.03]), *one-sided* $t(105) = -2.29$, $p = .012$, $d = -0.45$, meaning that the favorite was judged to have more efficacy than the underdog, which makes sense given that participants only know the players' final ranks. In the *both moving* condition ($M = 1.03$, 95% CI [0.73, 1.32]), the difference is greater than zero, *one-sided* $t(102) = 6.98$, $p < .001$, $d = 1.38$, signaling that the underdog seems more efficacious than the favorite, and greater than in the *both static* condition, $t(207) = 6.97$, $p < .001$, $d = 0.97$. Participants in the *favorite falling* condition ($M = 0.81$, 95% CI [0.51, 1.12]) also rated the underdog's efficacy as higher than the favorite's and the difference is greater than zero, *one-sided* ts > 5.25, ps <.001, ds > 1.00. In the *underdog rising* condition, the difference was not statistically different from zero ($M = 0.09$, 95% CI [-0.13, 0.30]), *one-sided* $t(105) = 0.81$, $p = .210$, $d = 0.16$, indicating that the players seemed to have similar efficacy. The efficacy difference was no different in the *favorite falling* condition than in the *both moving* condition, $t(211) = 1.00$, $p = .319$, $d = 0.14$, but all other efficacy differences were different from one another, ts > 2.20, ps < .029, ds > 0.30. Thus, information about the underdog rising is enough to bring his efficacy measurement to equal with the favorite's, but information about the favorite falling lowers the efficacy of the favorite enough that he is actually judged to have less efficacy than the underdog.

Mediation. I tested three mediation models to understand the effect of condition on expectation to win. For simplicity, I only included the *both moving* and *both static* conditions in the mediation model (coded as 1 and 0, respectively). First, to test my primary prediction that momentum changes perceived efficacy of the players which then changes expectations of winning, I ran a two-step 10,000 sample bootstrapping mediation model (Preacher & Hayes, 2008); this model supported my prediction and revealed a statistically significant indirect effect, 95% bias-corrected CI = [0.76, 2.38]. Second, to test a simpler prediction that momentum alone mediates the effect of condition on expectation to win, I ran another 10,000 sample bootstrapping model including perceived momentum as the mediator; this model revealed another significant indirect effect, 95% bias-corrected CI = [1.59, 3.71]. Finally, I tested an alternative possibility that only efficacy mediates the relationship between condition and expectations; this model did not have a statistically significant indirect effect, 95% bias-corrected CI = [-0.02, 1.53], indicating that momentum remains an important piece of the equation.

Exploratory questions. See Appendix 2 for an analysis of the exploratory items.

Controls. I also tested whether the effect of condition on expectation to win and momentum changed when I included familiarity with tennis, knowledge of tennis, and history as a tennis player as covariates in the generalized linear model analysis; the patterns remained exactly the same with no changes in statistical significance. There were no effects of participants' own familiarity, knowledge, or playing history on their beliefs about who would win the match or their perceptions of momentum.

Discussion

These results show that any knowledge about movement in rankings is enough to change observers' perceptions of the players' momentum, efficacy, and likelihood of winning a match. Perhaps most interesting, the effect of momentum in this efficacious domain is strong enough that observers believed the lower-ranked underdog was actually more likely to win the match with knowledge about both players' trajectories, despite the favorite still being ranked higher. Furthermore, the effect of momentum on expectations to win seems driven by, or at least closely linked to, beliefs about the players' efficacy. This supports my broader prediction that momentum only seems to change beliefs about performance because it signals that the players have efficacy. However, one concern about these findings is that observers were not incentivized for accuracy; their reported expectations that the underdog would win might be driven more by what they want than what they actually believe. To test this, I incentivize accuracy in Experiment 2.

Experiment 2: Incentivized Predictions

To better understand the true effect of momentum and efficacy on expectations to win, in Experiment 2 I provided participants with a financial incentive if they correctly predicted the result of a competition. I further generalized my results in Experiment 1 by testing another competition domain: basketball instead of tennis.

Method

Participants. I again predetermined 100 participants in each of four experimental conditions. In total, I recruited 407 participants (158 female, 2 gender non-binary, M_{age} = 32.04 years, 95% CI [31.11, 32.98]) through Amazon Mechanical Turk who completed a survey in exchange for $0.40.

Experimental design. I had the same four conditions as in Experiment 1 but I used NCAA women's basketball teams—Mississippi State and Kentucky—instead of tennis players to make sure my effect was not specific to tennis players.

Procedure. The procedure was nearly identical to the procedure in Experiment 1. The Mississippi State team (the favorite) and the Kentucky team (the underdog) were selected because they were close in current rank, but had moved significantly in the past week (favorite had fallen from 11th to 14th, underdog had risen from 18th to 16th), and they were scheduled to play against each other. If participants correctly selected the winner of the game, they were awarded an additional $0.50. I collected the momentum and efficacy measures described in Experiment 1 for each team. I again combined the three efficacy items into a single index for each team for analysis (αs = .88 for the favorite team and .82 for the underdog team). I did not ask the exploratory questions in this experiment but I did control for familiarity with basketball by adapting the same questions I used in Experiment 1, modified to apply to basketball instead of tennis.

Results

Expectation to win. Again the momentum manipulation significantly affected expectation to win, $F(3, 403) = 3.37$, $p = .019$, $\eta^2 = 0.02$ (see Figure 3). However, unlike in Experiment 1, in both the *favorite falling* and the *underdog rising* conditions (*M*s = 0.34 and 0.30, 95% CIs [0.24, 0.43] and [0.21, 0.39], respectively), there were no differences in the underdog's expectation to win from the *both static* condition ($M = 0.27$, 95% CI [0.19, 0.36]), *t*s < 0.99, *p*s > .326, *d*s < 0.14, nor were they different from one another, $t(203) = 0.60$, $p = .555$, $d = 0.08$. But when participants had information about movement for both teams, in the *both moving* condition ($M = 0.47$, 95% CI [0.37, 0.57]), they were significantly more likely to pick the underdog to win than in the *both static* and *underdog rising* conditions, *t*s > 2.51, *p*s < .013, *d*s > 0.36, and marginally more likely than in the *favorite falling* condition, $t(195) = 1.90$, $p = .059$, $d = 0.27$. Thus perceived momentum still affected expectation to win in this experiment, but the effect seems to be somewhat attenuated by the monetary incentive compared to Experiment 1's effects. Whereas in Experiment 1, the expectation to win fully flipped to the underdog, in this experiment, in the *both static*, *favorite falling* and *underdog rising* conditions, participants expected the favorite to win, *one-sided t*s > 3.46, *p*s < .001, *d*s > 0.69, and they were torn between the underdog and the favorite in the *both moving* condition. The expectation to win value was no different from 0.5, which equally preferences the two players, *one-sided* $t(95) = 0.61$, $p = 0.27$, $d = 0.13$.

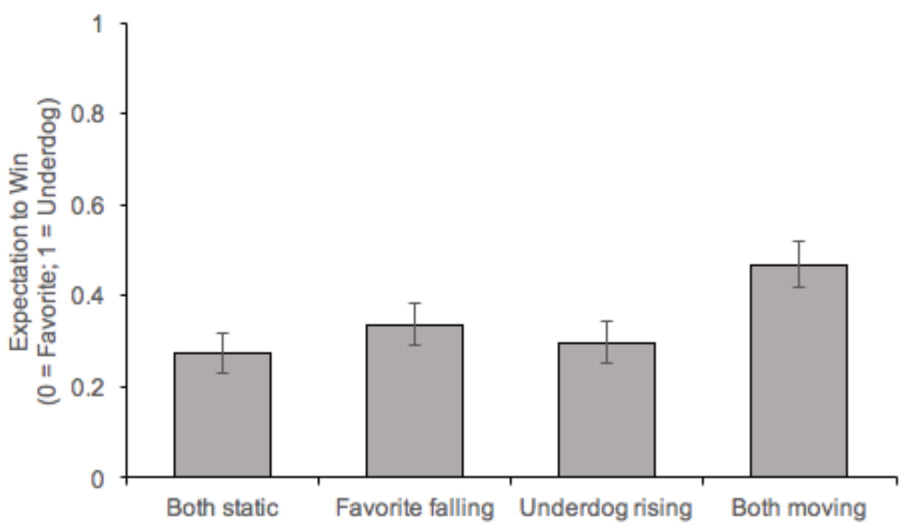


Figure 3. Expectation to win by experimental condition for Experiment 2. Error bars represent ± one standard error around the mean.

Perceived momentum. These changes in expectation were reflected in beliefs about the underdog's momentum (see Figure 4), $F(3, 403) = 9.55$, $p < .001$, $\eta^2 = 0.07$. In this experiment, the underdog was rated as having the least momentum in the *both static* condition ($M = 3.94$, 95% CI [3.51, 4.38]), and it was significantly lower than in any other condition, *t*s > 2.27, *p*s < .025, *d*s > 0.31. The momentum was judged to be the same in the *favorite falling* and *underdog rising* conditions (*M*s = 4.72 and 5.15, 95% CIs [4.19, 5.25] and [4.65, 5.66], respectively), *t*(203) = 1.17, *p* = .243, *d* = 0.16. In the *both moving* condition (*M* = 5.85, 95% CI [5.27, 6.44]), the underdog had significantly more momentum than in the *both static* and *favorite falling* conditions, *t*s > 2.84, *p*s < .005, *d*s > 0.40, and marginally more momentum than in the *underdog rising* condition, *t*(198) = 1.81, *p* = 0.072, *d* = 0.26.

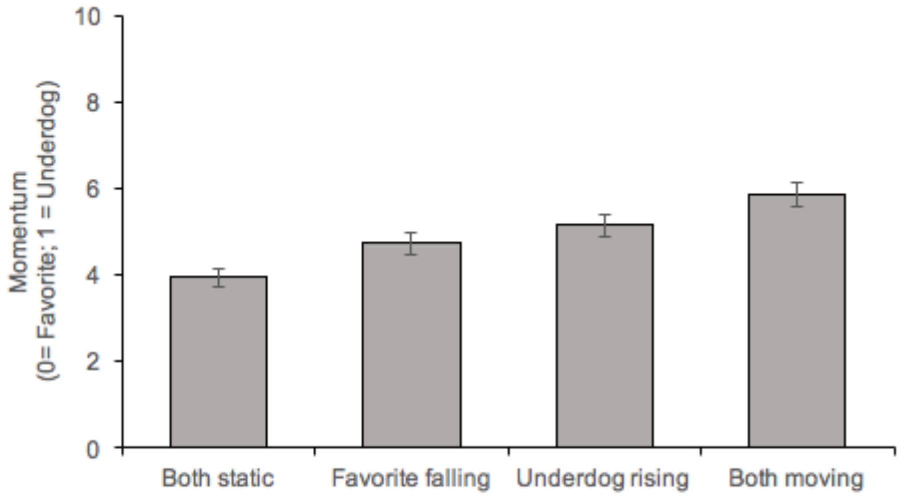


Figure 4. Perceived momentum by experimental condition for Experiment 2. Error bars represent ± one standard error around the mean.

Perceived efficacy. Efficacy showed similar results; condition affected perceptions of efficacy, *F*(3, 403) = 7.21, *p* < .001, η^2 = 0.05. As in Experiment 1, I report efficacy measures as a difference score in which positive indicates that the underdog has more efficacy. The efficacy measures are highly correlated between teams, *r*(405) = 0.15, *p* = .002, and the difference score is highly correlated with judgments of momentum, *r*(405) = 0.60, *p* < .001. In this case, the differences in efficacy mirror the judgments about expectation to win. In the *both static* condition (*M* = -0.67, 95% CI [-1.01, -0.32]) and the *underdog rising* condition (*M* = -0.36, 95% CI [-0.65, -0.06]), the difference is significantly less than zero, indicating that the favorite has more efficacy than the underdog, *one-sample t*s > 2.40, *p*s < .009, *d*s > 0.47, and these judgments are the same as one another, *t*(208) = 1.35, *p* = .179, *d* = 0.19. In the *favorite falling* condition (*M* = -0.19, 95% CI [-0.61, 0.22]), the difference is not significantly different from zero, indicating that the two teams have equal efficacy, *one-sample t*(100) = 0.92, *p* = 0.18, *d* = 0.18. However, this condition is no different from the *underdog rising* condition, *t*(203) = 0.66, *p* = .512, *d* = 0.09, and only marginally different from the *both static* condition, *t*(205) = 1.76, *p* = .080, *d* = 0.25. The *both moving* condition is different from all three of the other conditions (*M* = 0.58, 95% CI [0.09, 1.06]), *t*s > 2.40, *p*s < .018, *d*s > 0.34. When participants have information about both teams' movement, they think that the underdog has significantly more efficacy than the favorite, *one-sided t*(95) = 2.36, *p* = .010, *d* = 0.48. As with expectation to win, the largest gap between efficacy judgments occurs between the *both static* condition and the *both moving* condition and information about only one team's movement is not quite strong enough to change perceptions of efficacy.

Mediation. I again used only the both static and both moving conditions to test mediation. A two-step mediation model showed a significant indirect effect, supporting my primary prediction that momentum and efficacy sequentially mediated the effect of condition on performance expectations, 95% bias-correct CI = [0.44, 1.35]. Two simple mediation models indicated that momentum alone also mediated the effect of condition, 95% bias corrected CI = [0.95, 2.46], but efficacy alone was not a significant mediator, 95% bias-corrected CI = [-0.14, 0.68].

Controls. As in the first experiment, including the control questions about familiarity with NCAA women's basketball, following women's basketball, and playing basketball did not change the effect that condition had on either momentum or expectation to win.

Discussion

These results conceptually replicated my findings in Experiment 1, although the effects were attenuated by the financial incentive and observers were much less likely to vote for the underdog to win in Experiment 2 than Experiment 1. Critical for my prediction, however, perceived momentum still affected perceived efficacy and expectation of winning. Neither Experiment 1 nor Experiment 2 disentangles the effect of momentum from the effect of efficacy, however; these variables are highly correlated, $r(405) = 0.60$, $p < .001$. In Experiment 3, I test whether perceived momentum will still influence performance beliefs even when it signals nothing about players' efficacy.

Experiment 3: Momentum Without Efficacy

Experiments 1 and 2 suggest that efficacy mediates the effect of momentum perceptions on expectation to win, but to what extent does efficacy truly cause this effect, rather than merely being closely correlated with momentum perceptions? I predict that momentum only influences performance expectations *because* it seems to reflect players' efficacy. To test this, in Experiment 3 I tell participants that players' change in rankings is not due to their underlying efficacy, thereby disentangling momentum from efficacy.

Method

Participants. I predetermined a sample size of 100 participants for three experimental conditions. I recruited 307 participants (140 female, 4 gender non-binary, $M_{\text{age}} = 35.51$ years, 95% CI [34.13, 36.90]) through Amazon Mechanical Turk who completed a survey in exchange for $0.40.

Experimental design. I randomly assigned participants to one of three conditions: *no momentum*, *momentum with efficacy*, and *momentum without efficacy*. The first two conditions were the same as the *both static* and *both moving* conditions in Experiments 1 and 2, respectively. In the *momentum without efficacy* condition, I gave explanations other than efficacy for players' change in rankings.

Procedure. The procedure was the same as in Experiment 1 for the *no momentum* and *momentum with efficacy* conditions, using the same tennis players (Raonic, the favorite, and Simon, the underdog).

Momentum without efficacy. In the *momentum without efficacy* condition, participants read: "In the past few matches against lower-ranked opponents, Raonic was struggling with an ankle injury. He is currently completely healthy. Due to a strange set of coincidences, Simon's previous few higher-ranked opponents have either been playing injured or forfeited part-way through the match." This information was designed to provide an alternative explanation beyond efficacy for why each player was moving in the rankings.

Expectation to win. I asked who participants expected to win the match using the same question format as in the first two experiments.

Perceived efficacy and perceived momentum. Participants then answered the momentum and efficacy questions in a counterbalanced order to make sure the answer to one did not impact the answer to the other. In this experiment, to simplify my analysis, I only asked about perceived momentum and efficacy for the underdog. I measured momentum using a slider scale that was anchored at -10 (*Negative momentum*) and 10 (*Positive momentum*) and was initially set at 0. I again combined the three efficacy items into a single index for analysis ($\alpha = .89$).

Exploratory questions and controls. I also included the exploratory question about who the participants wanted to win and all the control questions described in Experiment 1.

Results

Expectation to win. As in the first two experiments, the momentum manipulation significantly impacted expectation to win, $F(2, 304) = 38.7$, $p < .001$, $\eta^2 = 0.20$, with a much higher expectation to win in the *momentum with efficacy* condition ($M = 0.81$, 95% CI [0.73, 0.89]) than in the *no momentum* condition ($M = 0.32$, 95% CI [0.23, 0.41]), $t(207) = 8.12$, $p < .001$, $d = 1.13$.

Perceived momentum. As in the previous experiments, these two conditions also had different ratings of momentum. In the *momentum with efficacy* condition ($M = 6.07$, 95% CI [5.44, 6.70]) the underdog was judged to have more momentum than in the *no momentum* condition ($M = 2.78$, 95% CI [2.09, 3.47]), $t(207) = 6.95$, $p < .001$, $d = 0.97$, which acts as a manipulation check.

Perceived efficacy. The underdog also seemed to have higher efficacy in the *momentum with efficacy* condition ($M = 8.43$, 95% CI [8.20, 8.67]) than in the *no momentum* condition ($M = 7.98$, 95% CI [7.71, 8.25]), $t(207) = 2.50$, $p = .013$, $d = 0.35$.

Mediation. The link between condition and expectation to win in the *no momentum* and *momentum with efficacy* conditions was mediated by momentum (95% bias corrected CI = [0.41, 1.21]), but not by efficacy (95% bias corrected CI = [-0.03, 0.16]). These results replicate the findings from my first two experiments.

Momentum without efficacy. I next examine perceptions of momentum and efficacy in the *momentum without efficacy* condition. Participants believed the underdog had more momentum in the *momentum without efficacy* condition ($M = 4.19$, 95% CI [3.53, 4.86]) than in the *no momentum* condition, $t(206) = 2.90$, $p = .004$, $d = 0.40$, but not as much as in the *momentum with efficacy* condition, $t(195) = 4.07$, $p < .001$, $d = 0.58$ (see Figure 5).

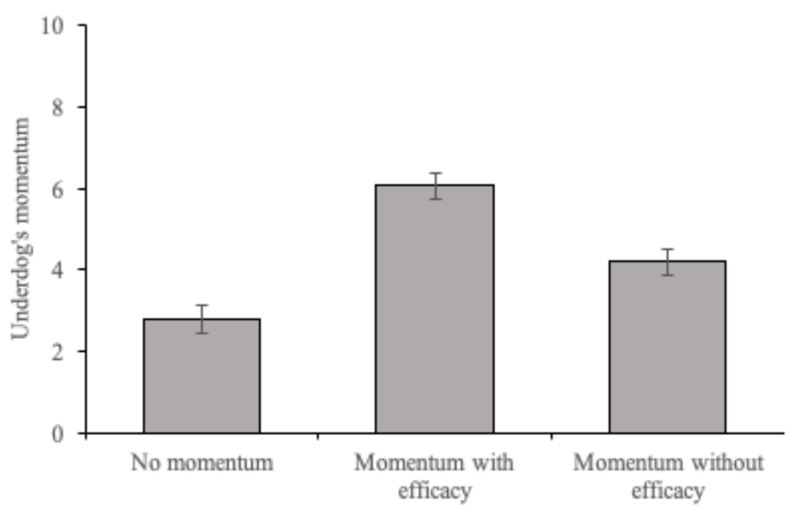


Figure 5. Underdog's momentum by experimental condition for Experiment 3. Error bars represent ± one standard error around the mean.

Participants rated the underdog's efficacy the same in the *momentum without efficacy* condition (M = 7.71, 95% CI [7.43, 7.99]) as in the *no momentum* condition, $t(206) = 1.38$, p = .170, d = 0.19, and less than in the *momentum with efficacy* condition, $t(195) = 3.93$, p > .001, d = 0.56. Participants who read information about the underdog's injured previous opponents and the favorite's previous injury knew that the players had been moving through the rankings, and even rated the underdog as having slightly more momentum than those who knew only the players' static ranks. Despite this knowledge, participants in the *momentum without efficacy* condition did not perceive the underdog to have higher efficacy than those in the *no momentum* condition (see Figure 6), signifying that the experience of momentum is deeply tied to efficacy. My manipulation completely erased the effect of movement on efficacy and attenuated it for momentum.

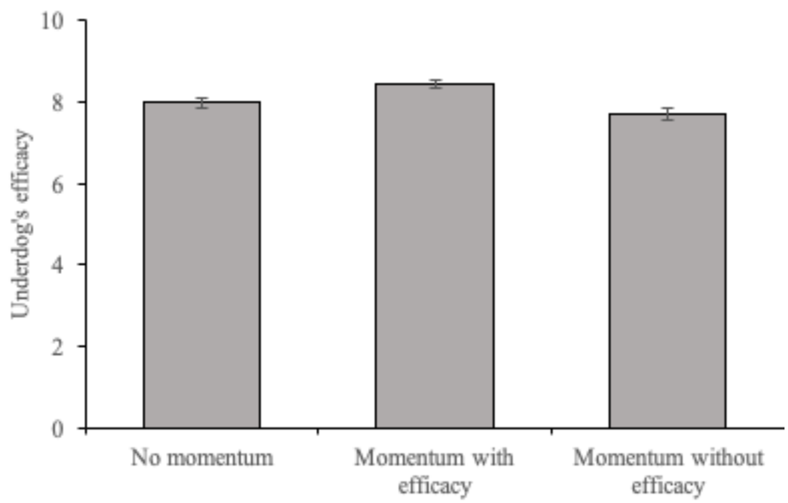


Figure 6. Underdog's efficacy by experimental condition for Experiment 3. Error bars represent ± one standard error around the mean.

Expectation to win. Most importantly, the efficacy manipulation entirely erased the effect of momentum on expectation to win, suggesting that there must be some signal of efficacy for that effect to occur. Expectation for the underdog to win in the *momentum without efficacy* condition (M = 0.34, 95% CI [0.24, 0.43]) was not significantly different from in the *no momentum* condition, $t(206) = 0.28$, p = .777, d = 0.04, and was significantly lower than in the *momentum with efficacy* condition, $t(195) = 7.57$, p < .001, d = 1.08 (see Figure 7).

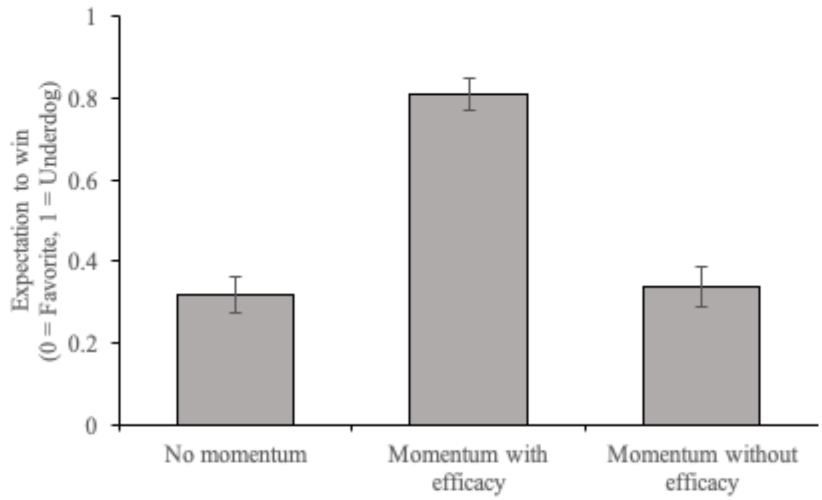


Figure 7. Expectation to win by experimental condition for Experiment 3. Error bars represent ± one standard error around the mean.

Mediation. I predicted that the difference in expectations of winning between the *momentum with efficacy* and *momentum without efficacy* conditions was due primarily to perceived efficacy, not to momentum. Indeed, in a 10,000 sample bootstrapping mediation model, efficacy mediated the difference in expectations to win between the *momentum without efficacy* and *momentum with efficacy* conditions, 95% bias corrected CI = [0.02, 0.36]. However, in a separate model, momentum still mediated the effect, 95% bias corrected CI = [0.17, 0.73], suggesting that momentum still partly accounts for predictions about future performance even when it diverges from perceived efficacy.

Exploratory questions. See Appendix 2 for an analysis of the exploratory items.

Discussion

This experiment successfully disentangled efficacy and momentum by creating a condition in which observers believed players had momentum but not efficacy. When observers believed an underdog player had momentum but not efficacy, their expectations about whether a streak would continue were no different from when they believed the underdog had no momentum at all. This suggests that a primary reason why momentum predicts performance expectations is because it creates a belief that the player has efficacy; without this belief, the effect of momentum on expectations disappeared.

Experiment 4: High-Efficacy versus Low-Efficacy Competitions

To further examine my hypothesis that momentum impacts performances predictions only when efficacy is present, I tested the effect of momentum in a prototypical domain in which performance seems to depend on players' efficacy (i.e., basketball) versus a domain in which performance seems to not depend on players' efficacy (instead relying on other forces like luck; i.e., roulette). I expected that gaining (versus losing) momentum would lead observers to predict a subsequent success in basketball but a subsequent failure in roulette, and these beliefs would be mediated by perceived efficacy and luck, respectively.

Method

Participants. I predetermined a sample size of 50 participants in each of four experimental conditions. In total, I recruited 208 participants (100 female, M_{age} = 37.25 years, 95% CI [35.54, 38.97]) through Amazon Mechanical Turk who completed a survey in exchange for $0.20.

Experimental design. I used a 2 (momentum: gaining or losing) × 2 (sport: high-efficacy or low-efficacy) between-participants design.

Procedure. I randomly assigned participants to see information about a basketball player (high efficacy) or a gambler (low efficacy). After they read a consent form and agreed to participate, I gave participants the following information: "In today's study you'll be making some predictions about a [gambler's/basketball player's] performance. This [gambler/basketball player] is named Joe. He's a professional [gambler/basketball player] who has been [competing annually in Vegas/playing in the NBA] for the past five years." I then told them to "imagine there's a roulette competition where Joe has bet on the color "red" six times in a row. Each time the wheel hits red, Joe earns $10,000, but each time the wheel hits black, Joe loses $10,000," or to "imagine there's a basketball game where Joe has had the opportunity so far to take six free-throw shots. Each time Joe has the opportunity to take a shot, he can win one point for his team if makes the basket but gets no points if he misses the basket." I then randomly assigned participants to either the *gaining momentum* or *losing momentum* condition. In the *gaining momentum* condition, I told participants that the roulette wheel had hit "red" the last six spins in a row or that Joe had completed his free throws in all six previous attempts. In the *losing momentum* condition, I told participants that the roulette wheel had hit "black" the last six spins in a row or that Joe had missed his last six free throws.

Expectation to win. For my primary prediction measure, I asked participants to guess the outcome of the seventh event: "Do you think that the wheel is likely to land on red (thereby winning Joe $10,000) or to land on black (thereby losing Joe $10,000)?" (*It will land on red (Joe wins)*, *It will land on black (Joe loses)*) or "Do you think that Joe is likely to make this shot, or to miss it?" (*Joe will make the shot*, *Joe will miss the shot*).

Perceived momentum. To measure perceived momentum I asked, "How much negative or positive momentum would you say that Joe has?" using a slider scale that was anchored at -50 (*Very negative momentum*) and 50 (*Very positive momentum*) and was initially set at 0 (*Neither negative nor positive momentum*).

Perceived efficacy. To measure perceived efficacy, participants then completed three measures about Joe: "In general, how **skilled** of a player do you think Joe is?", "In general, how much **trust** do you have in Joe to perform well?", "In general, how **confident** are you that Joe will perform well?" (1 = *not at all*, 7 = *very*; α = .93). I collapsed these three questions into one efficacy index for analysis.

Manipulation checks. As manipulation checks, at the end of the survey I asked, "To what extent does [making a free throw/winning at roulette] require skill?" and "To what extent does [making a free throw/winning at roulette] require luck?" (1 = *not at all*, 7 = *very much*).

Results

Manipulation checks. First, the manipulation checks confirmed that participants perceived that basketball required more skill (M = 6.03, 95% CI [5.78, 6.28]) than roulette (M = 2.50, 95% CI [2.19, 2.80]), $t(206) = 17.62$, $p < .001$, $d = 2.46$, but roulette required more luck (M = 6.13, 95% CI [5.89, 6.37]) than basketball (M = 2.90, 95% CI [2.62, 3.19]), $t(206) = 17.23$, $p < .001$, $d = 2.40$.

Expectation to win. I found my predicted cross-over interaction of game type and momentum on expectation to win, $F(1, 204) = 52.08$, $p < .001$, $\eta^2 = 0.20$, such that gaining momentum positively predicted expectations in the *high-efficacy* game, $t(103) = 7.78$, $p < .001$, d

= 1.53 ($M_{gaining}$ = 0.89, 95% CI [0.80, 0.97], M_{losing} = 0.29, 95% CI [0.16, 0.42]), but negatively predicted expectations in the *low-efficacy* game, $t(101) = -2.97$, $p = .004$, $d = -0.59$ ($M_{gaining}$ = 0.40, 95% CI [0.27, 0.54], M_{losing} = 0.69, 95% CI [0.55, 0.82]; see Figure 8). There was also a main effect of momentum condition on expectation, $F(1, 204) = 7.06$, $p = .008$, $\eta^2 = 0.03$, but no main effect of game type on expectation to win, $F(1, 204) = 0.59$, $p = .444$, $\eta^2 < 0.01$.

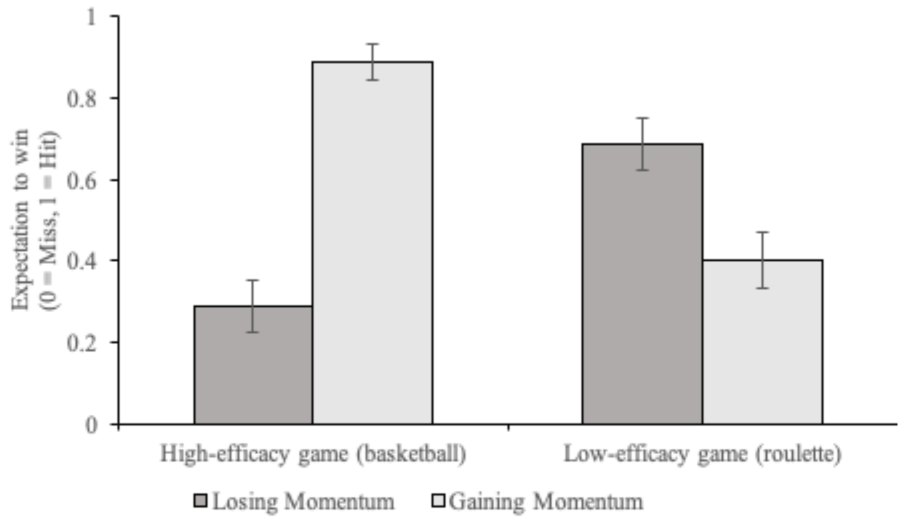


Figure 8. Expectation for Joe to make the next basket (high-efficacy game) or hit red on the next spin of the roulette wheel (low-efficacy game) by momentum condition and game type in Experiment 4. Error bars represent ± one standard error around the mean.

Perceived momentum. I found a main effect of momentum condition on perceptions of momentum, $F(1, 204) = 256.16$, $p < .001$, $\eta^2 = 0.53$, signaling that my manipulation affected perceived momentum as intended. There was no main effect of game type ratings of momentum, $F(1, 204) = 0.25$, $p = .616$, $\eta^2 < 0.01$, but there was an interaction, $F(1, 204) = 15.39$, $p < .001$, $\eta^2 = 0.05$, such that the effect of condition on momentum was larger in the *high-efficacy* game, $t(103) = 16.97$, $p < .001$, $d = 3.34$ ($M_{gaining}$ = 34.92, 95% CI [30.82, 39.03], M_{losing} = -29.02, 95% CI [-35.41, -22.63]), than in the *low-efficacy* game, $t(101) = 6.92$, $p < .001$, $d = 1.38$ ($M_{gaining}$ = 18.19, 95% CI [12.23, 24.11], M_{losing} = -15.06, 95% CI [-22.71, -7.41]; see Figure 9), likely because efficacy and momentum are so closely related.

Figure 9. Joe's perceived momentum by momentum condition and game type in Experiment 4. Error bars represent ± one standard error around the mean.

Perceived efficacy. I also analyzed how momentum condition and game type affected ratings of efficacy. I found my predicted interaction, $F(1, 204) = 35.81$, $p < .001$, $\eta^2 = 0.11$ (see Figure 10), such that the momentum condition had no effect on perceived efficacy of the *low-efficacy* player ($M_{\text{gaining}} = 4.15$, 95% CI [3.71, 4.59], $M_{\text{losing}} = 3.73$, 95% CI [3.33, 4.13]), $t(101) = 1.40$, $p = .166$, $d = 0.28$, but a significant effect on the perceived efficacy of the *high-efficacy* player, $t(103) = 10.81$, $p < .001$, $d = 2.13$, in that the *high-efficacy* player gaining momentum seemed to have more efficacy ($M = 5.93$, 95% CI [5.68, 6.18]) than the one losing momentum ($M = 3.17$, 95% CI [2.72, 3.62]). There were also significant main effects for both momentum condition, $F(1, 204) = 66.65$, $p < .001$, $\eta^2 = 0.21$, and game type, $F(1, 204) = 10.14$, $p = .002$, $\eta^2 = 0.03$.

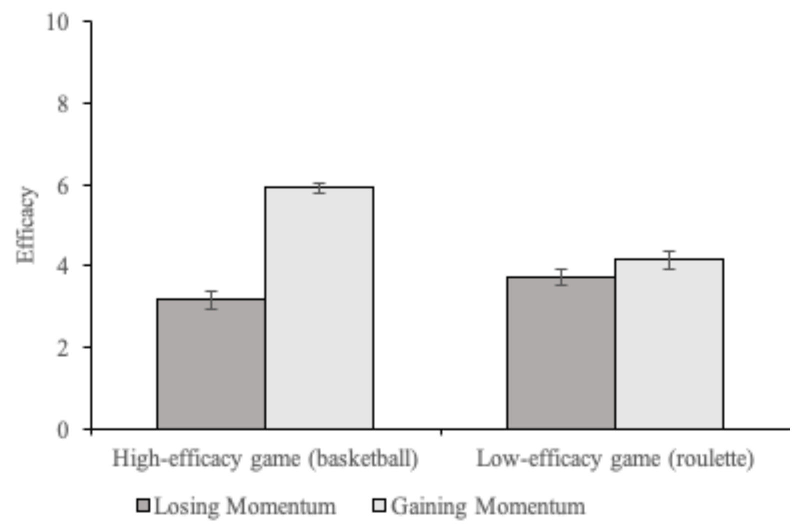


Figure 10. Joe's perceived efficacy by momentum condition and game type in Experiment 4. Error bars represent ± one standard error around the mean.

Streak reversal. I found an effect of game type on both whether Joe's luck had run out ($M_{\text{high-efficacy}} = 2.51$, 95% CI [2.08, 2.94], $M_{\text{low-efficacy}} = 4.21$, 95% CI [3.68, 4.74]) (in the *gaining* *momentum* condition) and whether Joe was due for a win ($M_{\text{high-efficacy}}$ = 3.17, 95% CI [2.64, 3.70], $M_{\text{low-efficacy}}$ = 4.20, 95% CI [3.70, 4.69]) (in the *losing momentum* condition). Both of those questions ask whether participants think the streak will reverse, and as predicted, participants in the *low-efficacy* condition, which was rated as more luck-based and less skill-based, are more likely to think the streak will reverse than those in the *high-efficacy* condition, $F_{\text{gaining}}(1, 103) = 25.11$, $p < .001$, $\eta^2 = 0.20$ and $F_{\text{losing}}(1, 103) = 7.97$, $p = .006$, $\eta^2 = 0.07$. This supports my prediction that efficacy is necessary to believe a streak of successes will continue.

Discussion

As predicted, participants viewed roulette as a low-efficacy game and basketball as a high-efficacy game, leading to some interesting patterns in expectation to win, momentum and efficacy. Supporting my primary hypothesis, this experiment showed that efficacy is an important component in the link between momentum and increased performance expectations. In basketball, an activity in which performers were rated to have substantial efficacy, participants thought a streak would continue. In roulette, however, an activity in which performers were rated to have very little efficacy, participants thought the player's streak would reverse. Momentum and efficacy also operate differently depending on the type of activity; in both perceived efficacy and perceived momentum, high-efficacy games see a greater swing from the *gaining momentum* condition to the *losing momentum* condition than low-efficacy games. I also show that participants are more likely to attribute a reversal of streak in a low-efficacy game to either luck running out or being due for a win, two attributions that are not efficacious.

Discussion (Chapter 2)

Momentum is a powerful psychological force in competitions that can change expectations of winning. However, four experiments provide evidence that an increased expectation to win based on an actor's gaining momentum is contingent on the attribution that the momentum signals efficacy. When momentum appears not due to a player's efficacy but instead to external forces (e.g., luck, injury) it does not change expectations of winning or even reverses expectations. These results are robust to different ways of manipulating momentum (e.g., information about ranking changes, observing a streak) and different competition domains.

Theoretical Contribution

This research makes many theoretical contributions. First, I integrate two prior literatures that use the term "momentum" differently and show completely opposing results—the "hot hand" and the gambler's fallacy. The "hot hand" predicts that a streak of successes will lead to continued successes, whereas the gambler's fallacy predicts that a streak of successes will lead to a reversal of fortune. To my knowledge, only a few papers have tried to reconcile these opposing literatures. These papers claim that the difference between the opposing expectations lies in whether the task involves people (Ayton & Fischer, 2004; Croson & Sundai, 2005) or randomness (Burns & Corpus, 2004). However, I propose and show that the extent to which momentum signals efficacy is a single mechanism that can parsimoniously explain the different beliefs. When the initial success seems to require efficacy, people predict that people gaining momentum will continue to experience success (Experiments 1–4). But when the initial success does not seem to involve efficacy, people predict that those gaining momentum will not have future success over and above those losing momentum, or may even reverse their current streak (Experiments 3 and 4).

Within the domain of competitions, I advance prior research by more clearly explicating how momentum influences performance expectations: by signaling efficacy. Most prior research on momentum is plagued by the problem that momentum and skill go hand-in-hand. That is, the player who seems to have momentum is often also the player that truly has more skill, so it is not clear whether observers believe the momentum itself causes greater performance or whether they believe skill causes greater performance. The latter belief (that the more skilled person will win a contest) is entirely normative, so it is particularly important to differentiate between these two explanations. I am among a small handful of researchers who try to disentangle momentum and efficacy. I found that in low-efficacy domains such as games of luck, observers may perceive momentum but not efficacy, and this form of momentum does not increase expectations of winning (and instead actually reduces expectations) (Experiment 4). I further found that even in high-efficacy performance domains, it is possible to remove the attribution of efficacy from a momentum experience. Unlike momentum that signals efficacy, momentum without efficacy appears not to affect performance beliefs (Experiment 3).

Limitations and Future Directions

This research elicits several possible directions for future work. For one, this research does not explore the boundaries of perceived momentum in changing performance expectations because it only uses very closely ranked opponents. Participants in Experiment 3 seemed to extrapolate linearly from an increase in rankings, so if two players are too far apart initially, momentum is likely not enough to counteract the initial gap and change expectations of winning. For instance, consider a player ranked 40th who had improved from 50th. If she faces a player ranked 30th who had dropped from 25th, even though the lower-ranked player moved twice as far in the rankings as her opponent, participants might be unlikely to expect even her efficacious momentum to change her binary performance expectations. Perhaps, however, observers would expect that player's losing score to be better than a similarly ranked player who had only risen five spots in the rankings. To answer that question, future research should explore non-binary performance outcomes in high-efficacy domains.

This research also does not address the time scale element of momentum. Is the effect stronger when momentum occurs within a single game because it is most proximal, or when it builds from a set of games because it has persisted? What about momentum created over weeks and months, like that of a college football team, or even over years like that of a rivalry? Can longer-lasting momentum create the necessary change in efficacy to see an impact on performance expectations? More research is needed to uncover various boundaries to the effect of momentum on perceived performance.

It would also be interesting to determine whether observers perceive efficacy in animals or machines. For instance, would people predict that computers like Deep Blue (designed to beat chess grand masters) or Watson (designed to win Jeopardy!) would continue a streak of successes against an evenly skilled opponent? What about a race horse like Secretariat? Chess, Jeopardy!, and horse racing are all goal-driven competitions with a clear winner, but the actors I just mentioned are not sentient in the same way humans are and presumably lack constructs such as confidence or trust in their skills. However, both Deep Blue and Watson are constantly learning, so in some sense, their efficacy does change as they compete. Further research could shed more light on whether non-humans can have efficacy and how that type of efficacy affects predicted performance outcomes.

Finally, given that efficacy clearly plays an important role in performance expectations, it would be valuable to understand how efficacy is created and whether people intuitively understand how to maximize their own efficacy. I explore these and related questions about the provenance and consequences of efficacy in Chapter 3.

CHAPTER 3:

Eat That Frog, Even If You Don't Want To:
Predicted and Actual Effects of Task-Ordering on Efficacy

Abstract (Chapter 3)

When people sit down to tackle day-to-day tasks, how do they prioritize? Do they prefer to start with the easiest task and work up to the hardest one, or get the hardest one out of the way and coast through the easiest one? Do people believe completing tasks in increasing- or decreasing-difficulty order generates more efficacy, and which one actually does? Over nine total experiments in three parts, I answer these questions by examining how task ordering influences perceived and actual efficacy. In Part 1, Experiments 1–3 (N = 496) show that participants believe completing tasks in increasing-difficulty order will lead to greater felt efficacy relative to completing them in decreasing-difficulty order and thus prefer to complete tasks in increasing-difficulty order. In Part 2, Experiments 4–6 (N = 1232) find that completing tasks in decreasing-difficulty order creates more reported self-efficacy than completing them in increasing-difficulty order while Experiments 7–8 (N = 499) show no difference in task ordering on efficacy. All five experiments in Part 2 expose participants' beliefs from Part 1 as mispredictions to varying degrees. In Part 3 (N = 502), I am able to help people correct their predictions, and I find evidence that reporting efficacy multiple times moderates whether task ordering changes feelings of efficacy. Together, these experiments illuminate how people think task ordering can affect efficacy, how people prefer to complete tasks, and how task ordering actually does affect efficacy, which has important implications for our day-to-day lives.

"If it's your job to eat a frog, it's best to do it first thing in the morning. And if it's your job to eat two frogs, it's best to eat the biggest one first."—Mark Twain

For almost any important activity, people want to build their efficacy; they want to feel more skilled, more confident, and more trust in themselves. Building self-efficacy is a key goal in people's day-to-day lives. To that end, it is important to understand whether certain situations increase or decrease felt efficacy. Take, for example, completing tasks of different difficulty levels. In everyday life, we often have to decide how to complete multiple tasks that may vary by difficulty. How would you prioritize them, and why?

You might choose to start with the easiest task first and work your way up to the harder tasks. Conversely, you might choose to start with the hardest task to get it out of the way. In the quote above, Mark Twain advocates for doing the latter, but there has been very little academic study on the benefits and costs of each approach. Does one order lead to more efficacy than the other? If so, is it the order people expect it to be?

In this research, I will examine predictions about whether completing tasks in increasing- or decreasing-difficulty order leads to more efficacy, preferences about task ordering, and whether completing tasks in increasing- or decreasing-difficulty order actually creates more felt efficacy in performers. I will also compare participants' predictions to performers' reported efficacy in order to determine whether general intuitions about task ordering match reality.

Defining Task Difficulty

I focus on task-difficulty in this research because it is a salient attribute in our daily lives and it is relevant to efficacy in that you often need to build skill in order to tackle more difficult tasks. Difficulty is a relative measure, depending on both the perceiver and the task. Because task difficulty is inherently subjective, in two of the experiments, I allow the participants to determine the difficulty of each task for themselves. In the rest of the experiments, I define and operationalize a task as more difficult than another task if it requires more time or effort to complete and/or has a lower likelihood of success. Tasks that require different levels of effort can have the same likelihood of completion. For example, consider a maze. Any maze is able to be completed if the performer tries hard enough to do so, but the more lines and dead ends a maze has, the more effort the actor must exert to complete it, and thus the more difficult it is. Conversely, tasks that require the same amount of effort can vary greatly in likelihood of success. Take, for example, a GRE analogy question. Every question simply requires reading all the options and choosing one, but the task can vary in how likely an actor is to answer correctly depending on how familiar or arcane the words in the analogy are. Finally, tasks can vary in both effort and likelihood of success. Crossword puzzles are a great example. They can be made more difficult by adding clues, which would require more effort, or making the clues harder, which would lower the likelihood of success, or both.

In regards to the perceiver, task difficulty is also relative. Perceived difficulty can depend on many individual characteristics including familiarity with the task, expertise in certain areas, and expectations. Interestingly, even the exact same task might be perceived as more or less difficult by a single actor, depending on whether her expectations about the task difficulty are upheld or violated (Moore & Healy, 2008) and the difficulty of the task that she performed immediately prior (Hancock, Williams, Manning, & Miyake, 1995). In this work, I will aim to maintain the same relative difficulty between all the tasks, independent of the performer.

There are at least three adjacent variables that can be disambiguated from the construct of difficulty. The first two are *urgency* and *importance*. Zhu, Yang, and Hsee (2018) define urgency as “the state that requires immediate responsiveness” (p. 1) and importance as “the state that involves significant outcomes” (p. 1). However, a more difficult task does not necessarily need to be more urgent or more important. In a daily to-do list, for instance, all your tasks may need to be completed by the end of the day, but the tasks do not have different deadlines within that time frame.

Furthermore, task difficulty may at times align with unpleasantness. After all, when tasks are unpleasant to complete they typically feel more difficult (Steel, 2007). I am not interested in how aversive a task is, per se, though I do believe that can be a component or a subset of difficulty. In my experiments, I will control for task unpleasantness when manipulating difficulty, specifically selecting tasks that are similarly enjoyable but differ in difficulty of completion.

Predictions About Building Efficacy via Task Ordering

As previewed above, it is important to understand people’s predictions about how a specific ordering will affect their efficacy, which may further affect their preferences for ordering tasks. As we go about our day-to-day lives, we make choices about which tasks to complete and in which order, and these choices are largely driven by our personal preferences. I hypothesize that, generally, people will believe that completing tasks in increasing-difficulty order will increase their perceived efficacy, which may also affect their preferences for task ordering. There are four literatures that inform these predictions.

First, I suspect that people may have some lay beliefs that confidence must “build” with repeated successes (for an idea like this, see Shaw, Dzewaltowski, & McElroy, 1992). Although empirical data does not clearly indicate how confidence can be built in such a way, scholars such as Feltz and Weiss (1982) have suggested that efficacy can be improved through “gradual increases in skill improvement” brought on by “progressive activities” (p. 24). This lends support to the idea that completing tasks in increasing-difficulty order may lead to greater feelings of efficacy.

The learning literature makes a similar prediction. When people learn, they necessarily must start with easier tasks and work their way towards the harder ones (Mowrer, 1960). The easier tasks act as building blocks and the harder tasks often cannot be completed without sufficient knowledge. For instance, in learning math, you must understand how to count before you can add, and you must understand how to add before you can multiply. Because so much of people’s early lives are spent learning in this capacity, starting with the easy tasks and working their way up may be a more familiar order for them to pursue and could contribute to the belief that increasing-difficulty order creates more efficacy as well.

There is also evidence that people simply prefer to start with their easiest tasks. In a meta-analysis of the procrastination literature, Steel (2007) finds that task aversion and task delay are strong and consistent predictors of procrastination. Although task aversion is not necessarily the same as task difficulty (see discussion above), these findings suggest that people may prefer to start with their easier tasks and put off their harder tasks until later. Relatedly, the temporal discounting literature suggests that people may prefer to do hedonic tasks first, or even draw out their most favorable tasks in order to savor the experience (Frederick, Loewenstein, & O’Donoghue, 2002; Harris, 2012). Hedonic tasks are not necessarily easier (again, see discussion above) but these results indicate that people may prefer to do easier tasks first, especially if they tend to think of them as being more pleasant.

However, not all of the prior research findings support a preference for completing tasks in increasing-difficulty order. For example, Harris (2012) finds that some participants prefer to start with the hardest or most aversive task to avoid having to dread it, and Loewenstein (1987) proposes that people may choose to wait for pleasurable events (i.e., savoring) or speed up negative events, which would suggest that people may prefer to begin with their hardest tasks and save their easiest tasks until the end. Speeding through negative tasks instead of prolonging them does make people feel better about the experience (Nelson & Meyvis, 2008), and a number experiments do find that people express a preference for sequences that improve over time (Loewenstein & Prelec, 1993; Loewenstein & Sicherman, 1991; Varey & Kahneman, 1992).

Though people's general preferences for task ordering could be determined by many different factors, when the goal is to maximize efficacy, I hypothesize that people will prefer to complete tasks in increasing-difficulty order. Indeed, "Eat that Frog!," a popular motivational theory based off of Mark Twain's quote, advocates for starting with the hardest task of the day in order to maximize productivity and daily successes (Tracy, 2011). In this research I aim to provide experimental evidence that supports this theory.

The Actual Effect of Task Ordering on Efficacy

Although I believe people will predict that completing tasks in increasing-difficulty order will increase their efficacy more than decreasing-difficulty order, I think the opposite is actually true. As discussed in Chapter 1, prior research shows that a downward trajectory or a string of continued successes can elicit the perception of gaining momentum (Iso-Ahola & Dotson, 2014; Markman & Guenther, 2007; Shaw, Dzewaltowski, & McElroy, 1992). For instance, Silva, Cornelius, and Finch (1992) found that participants in a loss-win-win condition reported feeling more momentum than those in win-loss-loss condition. Further, Vallerand, Colavecchio, and Pelletier (1988) show that participants rated players as having more momentum when they had won five out of ten tennis games in a pattern of three losses, then one win, then two losses, then four wins (e.g., "0001001111") than those who had won the same number of games in a pattern of one loss, one win, two losses, two wins, one loss, one win, one loss, one win (e.g "0100110101"). If completing difficult tasks can create similar emotions to experiencing a loss, then participants who complete difficult tasks first will feel more losses early in the sequence and more wins as they continue through it. Therefore, the order of decreasing-difficulty creates a sense of momentum. Because efficacy and momentum are highly correlated (Attali, 2013; Cornelius et al., 1997; Gilovich, Vallone, & Tversky, 1985), I further predict that completing tasks in decreasing-difficulty order will not just make the performer feel momentum but will also build efficacy. With each subsequently easier task, participants will experience a stronger sense of confidence, skill, and trust in their abilities.

Why do people mispredict? These two hypotheses—that people believe completing tasks in increasing-difficulty order will increase efficacy but in reality the opposite is true—expose a misprediction. One reason people may make this misprediction is because they fail to accurately project the entirety of the experience.

In order to predict a future affective state or experience, people must employ mental simulations, but the simulations are often "mere cardboard cutouts of reality" (Gilbert & Wilson, 2007, p. 1354). These static representations cannot fully capture the phenomenological nature of the real-time experiences they are attempting to predict. Indeed, people often misunderstand how their own experiences and preferences build and grow over time (e.g., Kahneman & Snell, 1992; Kardas & O'Brien, 2018; Klein & O'Brien, 2018), suggesting that this misprediction is not unique. I propose that the more closely the prediction experience matches the actual experience, the more accurately participants will predict their future efficacy, since their mental simulations will hew closer to reality.

A consequence of efficacy: motivation. I also hypothesize that a greater sense of efficacy will lead to greater motivation. Efficacy influences such outcomes as persistence, effort, and task initiation (Bandura, 1977; Bandura, Adams, & Beyer, 1977), which are all components of motivation, and some work has also explored the link between efficacy and motivation directly (Bandura, 1993; Schunk, 1991; Schunk, 1995). Schunk reports a strong relationship between self-efficacy and motivation, proposes a model that links the two (1995), and even explicitly states that "heightened self-efficacy sustains motivation" (1991, p. 212). Similarly, Bandura (1993) claims that "self-efficacy beliefs contribute to motivation in several ways: They determine the goals people set for themselves; how much effort they expend; how long they persevere in the face of difficulties; and their resilience to failures" (p. 131).

Motivation, which Touré-Tillery and Fishbach (2014) define as "the psychological force that enables action," consists of two components, intrinsic and extrinsic. Intrinsic motivation is an inherent tendency to seek out challenges, a natural inclination towards exploration and mastery. Extrinsic motivation, on the other hand, occurs when one performs an activity in order to obtain a specific and separable outcome (Ryan & Deci, 2000). I argue that ordering tasks in decreasing- (vs. increasing-) difficulty will improve intrinsic motivation but will not affect extrinsic motivation, since task ordering does not change anything related to the ultimate outcome. Task ordering could, however, change how challenging participants perceive tasks to be and how difficult or easy it feels to master those tasks, which relates directly to intrinsic motivation. Because of the well-established link between efficacy and persistence (Barling & Beattie, 1983; Bouffard-Bouchard, 1990; Feltz & Lirgg, 1998; Moritz, Feltz, Fahrbach, & Mack, 2000; Stajkovic & Luthans, 1998; Walker, Greene, & Mansell, 2006), I test whether participants who complete tasks in decreasing-difficulty (vs. increasing-difficulty) order are more willing to complete another task in the same domain. I also collect a measure of state motivation, which indicates how generally motivated people feel after completing the tasks in one order or another.

Proposed Theory and Hypotheses

Based on the literature I have discussed so far, I believe that people will predict that increasing-difficulty order will promote greater efficacy and motivation than decreasing-difficulty order and, relatedly, prefer to complete tasks in increasing-difficulty order. However, I predict that completing tasks in decreasing-difficulty order will actually improve efficacy, motivation, and willingness to continue, contradicting general intuition.

Overview of Studies

This chapter contains three parts. Part 1 investigates people's predictions about how completing tasks in different orders might affect their efficacy preferences and for task ordering. Part 1 includes three experiments, each of which explores predictions and preferences using a different task. I find that people expect that starting with their easiest task and finishing with their most difficult will lead to greater feelings of efficacy than the opposite order and consequently prefer to complete tasks in that order. I also find that expected efficacy predicts order preferences. In Part 2, different participants actually complete the tasks that people made predictions about in Part 1 to test the accuracy of those predictions. In the first three experiments of Part 2, participants report that completing tasks in decreasing-difficulty order leads to more felt efficacy than completing those same tasks in increasing-difficulty order, but I find no effect on motivation or willingness to continue. In the final two experiments of Part 2, participants report no difference in efficacy by task order. I propose the number of times participants report their efficacy throughout the experiment as a potential moderator which could explain the difference in efficacy findings in the second part. Comparing the findings from Part 1 and Part 2 also reveals a misprediction. Participants consistently predict that completing tasks in increasing-difficulty order will lead to greater felt efficacy, while in fact I find either no effect of task ordering on efficacy or that completing tasks in decreasing-difficulty order creates more felt efficacy. In Part 3 I attempt to correct participants' mispredictions and to learn more about the mechanisms behind both the predictions and the order effect by incrementally changing the procedure for making predictions. In Part 3 I find that if participants predict their efficacy after observing and therefore simulating each round, they are correctly able to predict that completing tasks in decreasing-difficulty order leads to higher reported efficacy than completing them in increasing-difficulty order, while those predicting efficacy only once do not predict an effect of task order on efficacy. This finding also serves as supporting evidence for my proposed moderator from Part 2.

Part 1: Preferences and Predictions

In the first three experiments, people indicate a preference for ordering tasks, specifically focusing on how efficacious they think each order will make them feel. Participants also predict how much efficacy they would feel after completing the tasks in each order. In order to demonstrate the robustness of this effect, participants make these predictions about three different sets of tasks and in two distinct paradigms.

Experiment 1: Word Find Preferences and Predictions

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/yadks).

Participants. I predetermined a sample size of 200 within-subjects participants. I recruited 202 participants (93 female, M_{age} = 35.00 years, 95% CI [33.41, 36.59]) through Amazon Mechanical Turk who completed a survey in exchange for $0.40.

Procedure. I asked participants to imagine they were participating in a different MTurk survey. The task was to find words from a set of twelve letters. Participants saw the full instructions shown to participants who had previously completed the actual task. They were told to imagine they would "see twelve letters" and then have "one and a half minutes to write down as many 4+ letter words as possible using those letters." The word-find task had four rules: 1) each word could only be submitted once; 2) each letter could only be used once per word; 3) each word had to be at least four letters long; and 4) the words had to be real words that could be found in a dictionary. Participants' goal was to find as many words as possible following the four rules in the time allotted for each practice round and competition. For example, if their letter string was "*XHWYNEAJRTMF*," they could write words such as: "*near,*" "*meat,*" "*fret,*" "*wart,*" "*wharf,*" and so on. We told participants that previous participants completed three practice rounds of the task before a competition. We then told participants how difficult each of the sets of letters is—in the easy round participants found an average of 15.1 words, in the medium round participants found an average of 10.6 words, and in the hard round participants found an average of 8.1 words. We then told them that, if they were completing this task, they would have a competition after those three practice rounds and would earn a bonus if they beat their average score. Participants then gave their preferences for task ordering, followed by momentum and efficacy predictions. This is a within-subjects design, so each participant gave predictions about momentum and efficacy for both the increasing-difficulty order and the decreasing-difficulty order, counterbalanced. Participants then saw the following instructions: "Now imagine that you are assigned to see the practice rounds in the following order: **Easy round, then medium round, then hard round [Hard round, then medium round, then easy round]**.

Efficacy predictions. Participants answered three questions about predicted efficacy: "If you were assigned to see the practice rounds from **easy to medium to hard [hard to medium to easy]**, how skilled do you think you would feel at finding words, just before you entered the competition round?", "If you were assigned to see the practice rounds from **easy to medium to hard [hard to medium to easy]**, how confident do you think you would feel about finding words, just before you entered the competition round?", "If you were assigned to see the practice rounds from **easy to medium to hard [hard to medium to easy]**, how much would you trust your ability to find words, just before you entered the competition round?" (1 = *not at all [skilled/confident]*, 10 = *very skilled [confident/much]*; $\alpha = .93$ for increasing-difficulty and .94 for decreasing-difficulty). I collapsed these three questions into one efficacy index per condition for analysis.

Order preference. To measure participants' order preference, specifically with regards to the best way to build efficacy, I told them "Your goal is to feel the most **confident** and the most **skillful** before you start the competition" and then asked "To achieve this goal, how would you prefer to see the practice rounds?" (*Easy, then medium, then hard*, *Hard, then medium, then easy*, *It doesn't matter to me*). For exploratory analysis I also asked "Why did you choose [choice]?"

Momentum predictions. Although not the primary focus of the research, to be thorough, participants also answered the question "If you were assigned to see the practice rounds from **easy to medium to hard** [**hard to medium to easy]**, how much **momentum** do you think you would have entering the competition round?" Participants moved a slider that was anchored at -50 (*Negative momentum*) and 50 (*Positive momentum*) and was initially set at 0.

Performance predictions. As an exploratory measure, I also asked for participants' specific predictions about how many words they would find in each round.

Control variables. To control for participants' experience with the task, I asked "How familiar are you with word-find tasks similar to the ones you completed today?" (*I have never played a game like that before*, *I have played a game like that a few times*, *I sometimes play games like that*, *I frequently play games like that*, *I play games like that almost every day*). I also collected education, age, and gender.

Results

Efficacy predictions. Supporting my primary prediction, participants believed that completing tasks in increasing-difficulty order ($M = 7.25$, 95% CI [7.02, 7.47]) would create more efficacy than completing the tasks in decreasing-difficulty order ($M = 6.46$, 95% CI [6.18, 6.74]), $t(402) = 4.34$, $p < .001$, $d = 0.43$.

Preferences. Significantly more participants prefer to complete the tasks in increasing-difficulty order (53%) than decreasing-difficulty order (22%), $t(201) = 5.36$, $p < .001$, $d = 0.76$. The remaining 25% indicated that they were indifferent as to order. While 53% is not significantly more than half our participants, *one-sample* $t(201) = 0.84$, $p = .200$, $d = 0.12$, it is significantly greater than 33%, *one-sample* $t(201) = 5.67$, $p < .001$, $d = 0.80$, indicating that it was the most popular of the three choices.

Suggesting that predictions of efficacy drive people's preferences in this context, in a multiple logistic model excluding participants who were indifferent to task ordering, efficacy ratings predict order preference for both increasing-difficulty order ($\beta = -0.30$, $z = -2.58$, $p = .010$) and decreasing-difficulty order ($\beta = 0.59$, $z = 4.65$, $p < .001$), even when controlling for age, education, gender, and familiarity with the task, $\chi^2(5) = 11.96$, $p = .035$, and $\chi^2(5) = 35.59$, $p < .001$, respectively.

Other analyses. See Appendix 3 for analyses of all the other variables we collected, including momentum predictions and specific point value predictions.

Discussion

Participants believe that completing tasks in increasing-difficulty order will create more felt efficacy than completing them in decreasing-difficulty order and prefer to complete tasks in increasing-difficulty order. Ratings of efficacy predict order preference as well.

Experiment 2: Analogy Preferences and Predictions

In Experiment 2, I wanted to extend this finding to a different task both to confirm robustness and to have multiple predictions against which to test our actual findings. I also removed the indifferent option from the preference choice set to improve clarity. I chose analogies for this experiment because they are easy to vary in difficulty and should be familiar to most of our participants.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/se6tk/).

Participants. I predetermined a sample size of 200 within-subjects participants. I recruited 200 participants (94 female, M_{age} = 37.03 years, 95% CI [35.48, 38.57]) through Amazon Mechanical Turk who completed a survey in exchange for $0.50.

Task. The task in this experiment was to complete analogy questions. Each question was multiple choice and followed the format "____ is to ____ as____ is to____ ." For instance **catnip : cat :: bone : dog** can be read as "Catnip is to a cat as a bone is to a dog". For each analogy question, one of the four words was left blank, and participants selected one of four multiple-choice options to fill in the blank.

Procedure. The procedure was very similar to Experiment 1 with only a few minor changes. I described to participants how to complete an analogy and showed them a few examples. I then told participants "Each of the three rounds contains a different set of 6 analogies, which have been pre-tested and are either easy, medium, or hard difficulty. In the easy set, people correctly answered an average of 5.6 out of 6 analogies. In the medium set, people correctly answered an average of 3.6 out of 6 analogies. In the hard set, people correctly answered an average of 1.2 out of 6 analogies." I then showed participants the six analogies in each round for ten seconds, displaying the rounds in random order. After the instructions, participants answered the efficacy questions and the binary preference question as described in Experiment 1. I did not collect momentum measures or specific performance predictions in this experiment. All other measures were identical to those used in Experiment 1, with "answering analogies correctly" substituted for "finding words" where appropriate.

Results

Efficacy predictions. Replicating the findings from Experiment 1, participants again believed that completing tasks in increasing-difficulty order (M = 6.27, 95% CI [6.01, 6.57]) would create more efficacy than completing the tasks in decreasing-difficulty order (M = 5.75, 95% CI [5.43, 6.06]), $t(398) = 2.54$, $p = .011$, $d = 0.26$. None of the demographic or control variables predicted efficacy ratings.

Preferences. As in Experiment 1, significantly more participants preferred to complete the tasks in increasing-difficulty order (60%) than in decreasing-difficulty order (40%), $t(398) = 4.07$, $p < .001$, $d = 0.41$.

Efficacy ratings again predicted order preference in a multiple logistic regression model for both increasing-difficulty order ($\beta = -0.57$, $z = -5.68$, $p < .001$) and decreasing-difficulty order ($\beta = 0.64$, $z = 6.26$, $p < .001$), even when controlling for age, education, gender, and familiarity with the task, $\chi^2(5) = 49.03$, $p < .001$, and $\chi^2(5) = 66.07$, $p < .001$, respectively.

Discussion

These findings directly replicate the findings from Experiment 1 while extending into a slightly different domain and offering a clearer binary preference for completing tasks in increasing-difficulty order rather than decreasing-difficulty order. Participants further believe that completing tasks in increasing-difficulty order will create more felt efficacy than completing them in decreasing-difficulty order, and participants' efficacy ratings again predicted their order preferences.

Experiment 3: Job Application Preferences and Predictions

In this experiment, I used tasks related to applying for jobs rather than word puzzles to expand our knowledge about people's predictions and preferences around task ordering into a more externally valid set of tasks. This experiment also extends the findings from Experiments 1 and 2 by allowing participants to rate the difficulty of a number of tasks themselves and then to rank order them from easiest to hardest instead of manipulating difficulty using easy, medium, and hard rounds of tasks.[12]

Method

Participants. I recruited 94 first and second year students (68 female, 3 non-binary, M_{age} = 18.74 years, 95% CI [18.60, 18.89]) from the University of California, Berkeley's Experimental Social Science Laboratory (XLab) through SONA. I chose first and second year students because this was a pretest and I was planning to use third and fourth year students for the main experiment. I wanted to use a similar sample without removing participants from the pool I planned to use in the future.

Procedure. Because this experiment was designed to choose appropriate tasks related to applying for jobs for a future experiment, participants rated nine different tasks on the following measures: "How difficult this task is", "How pleasant this task is", "How important this task is in the job application process", "How much you want to complete this task" (1 = *not at all*, 10 = *very*). Participants also used a slider anchored at 0 minutes and 60 minutes with anchors every 10 minutes to indicate how long they thought the task would take to complete. As with the other prediction surveys, participants did not actually complete the tasks. For a detailed description of all nine tasks, please see Appendix 3.

Ranking. After evaluating all nine tasks in a random order, I asked participants to rank all nine tasks from easiest to hardest.

Efficacy predictions. As in Experiments 1 and 2, I asked participants to predict efficacy about both orders, counterbalanced. I told them "Now imagine that you are assigned to see the tasks in the following order: **Easy to hard [Hard to easy]**" and they answered the same efficacy questions as in Experiments 1 and 2 related to their skill, confidence, and trust in their abilities to apply for jobs.

Preferences. This preference question differs in an important way from the question I asked in Experiments 1 and 2. In this question, I do not explicitly reference efficacy as a motivator for their choice, so they may be basing these preferences on other factors. In this experiment I asked: "If you were forced to complete all nine tasks in either decreasing difficulty

[12] Experiment 3 is the only data in this chapter that is not pre-registered. I collected it as part of a pretest for Experiment 8 and included prediction and preference questions at the end to confirm that this new paradigm would replicate the findings from Experiments 1 and 2. The data were so compelling that I felt the need to include them despite their not being pre-registered.

order (starting with the hardest task and ending with the easiest) or increasing difficulty order (starting with the easiest task and ending with the hardest), which would you choose?" (*Easy to hard*, *Hard to easy*). They also answered the text prompt "Why did you choose [choice] as your preferred order?"

Likely to apply again. I also asked in each order "If you were assigned to see the tasks from **easy to hard [hard to easy]**, how likely do you think you would be to apply for another job after completing this application?" as an exploratory measure of motivation.

Control variables. Participants then answered "How familiar are you with applying for jobs?" (*I have never applied to a job before*, *I have applied to a few jobs*, *I have applied to many jobs*) and filled out their class year, major, whether they were currently applying for jobs, age, and gender.

Results

Efficacy predictions. Yet again, participants believed that completing tasks in increasing-difficulty order (M = 6.63, 95% CI [6.30, 6.96]) would create more efficacy than completing the tasks in decreasing-difficulty order (M = 6.02, 95% CI [5.67, 6.37]), $t(186) = 2.51$, $p = .013$, d = 0.37. I also ran a multiple regression model on efficacy ratings by condition, controlling for gender, age, year, whether they were applying for jobs currently, and their familiarity with applying for jobs, $F(6, 181) = 3.98$, $p < .001$, $R^2 = 0.12$. Condition significantly predicts ratings of efficacy ($\beta = -0.61$, $t = -2.59$, $p = .010$), indicating that participants believe completing tasks related to applying for jobs in increasing-difficulty order will lead to more efficacy than completing them in decreasing-difficulty order, even controlling for all the other variables in the model. Familiarity with applying for jobs is also a significant predictor in this model ($\beta = 0.64$, $t = 3.29$, $p = .001$).

Preferences. As in Experiments 1 and 2, significantly more participants prefer to complete the tasks in increasing-difficulty order (81%) than decreasing-difficulty order (19%), $t(186) = 10.69$, $p < .001$, $d = 1.57$. Expected efficacy predicts preferences in a multiple logistic regression when participants consider completing tasks in increasing-difficulty order ($\beta = -0.60$, $z = -2.76$, $p = .006$) even controlling for age, gender, year in school, whether they are currently applying for a job, and familiarity with applying for jobs, $\chi^2(6) = 15.35$, $p = .018$, but expected efficacy does not predict preferences when participants consider completing tasks in decreasing-difficulty order ($\beta = 0.25$, $z = 1.35$, $p = .176$), $\chi^2(6) = 8.05$, $p = .235$. This differs from Experiments 1 and 2, perhaps because efficacy was not an explicit component in the choice question or perhaps because so few people in this experiment ($n = 18$) prefer to complete tasks in decreasing-difficulty order.

Discussion

These results add to the robustness of the prediction and preference effects from Experiments 1 and 2 and improve the external validity of those findings. In Part 1, I present evidence from three different tasks and two different paradigms that people believe completing tasks in increasing-difficulty order will lead to feeling more efficacious than completing those same tasks in decreasing-difficulty order and prefer to complete tasks from easiest to hardest. I also find that participants' efficacy ratings predict their order preferences, which suggests that the preferences may be driven by a desire to maximize efficacy. The logical next step to determine whether these intuitions hold true is to have participants actually complete the tasks in these prediction experiments in increasing-difficulty order or decreasing-difficulty order and report their felt efficacy.

Part 2: Task Ordering and Efficacy in Action

In Part 1, I asked participants to predict how they would feel after completing tasks in different orders. In Part 2, I augmented the three tasks from Part 1 with two new tasks and asked participants to actually perform the tasks in either increasing-difficulty or decreasing-difficulty order and to report how much efficacy they feel. This allows me to compare predictions with reality and investigate whether the preferences people have for starting with their easiest task and finishing with their hardest makes sense given efficacy as a goal.

Experiment 4: Word Find Task

In order to test the validity of participants' predictions, I first present the results of an experiment in which participants completed the word find tasks described in Experiment 1. This data comes from Experiment 2 in Chapter 1 of this dissertation, which is pre-registered on the Open Science Framework (https://osf.io/wut82). For a full description of the methods, results and discussion, please see Chapter 1. For the purposes of this chapter, I will provide a brief overview of the methods and will only include efficacy-relevant results.

Method

Participants. I recruited 302 participants (174 female, 1 gender non-binary, M_{age} = 36.51 years, 95% CI [35.19, 37.83]) through Amazon Mechanical Turk who completed a survey in exchange for $0.70 with the opportunity to earn a substantial bonus.

Procedure. Experiment 4 employed two experimental conditions (between-subjects): *increasing-difficulty order* and *decreasing-difficulty order*. After signing a consent form and before beginning the word search task, participants read that they would "see twelve letters" and then have "one and a half minutes to write down as many 4+ letter words as possible using those letters." The word-find task had four rules: 1) each word could only be submitted once; 2) each letter could only be used once per word; 3) each word had to be at least four letters long; and 4) the words had to be real words that could be found in a dictionary. Participants' goal was to find as many words as possible following the four rules in the time allotted for each practice round and competition. For example, if their letter string was "XHWYNEAJRTMF," they could write words such as: "near," "meat," "fret," "wart," "wharf," and so on. I told participants that they would be completing three practice rounds of the task before a competition, and they responded to several attention check items to ensure that they had read and understood the rules. Because I am only interested in their efficacy ratings after their third practice round for the purposes of this experiment, the competition rules and pre-competition questions are not relevant here.

Perceived efficacy. To assess self-efficacy I asked three questions—"How skilled do you think you will be at finding words?", "How confident do you feel about finding words?", and "How much do you trust your ability to find words?" (1 = *not at all [confident/skilled/]*, 7 = *very [confident/skilled/much];* αs ≥ .97)—at five time-points throughout the experiment: before the first practice round, after each practice round, and after the competition. The dependent variable of interest is the efficacy rating collected after participants completed the third practice round.

Results

Perceived efficacy. Participants' self-efficacy was no different prior to completing the practice rounds in each condition ($M_{\text{decreasing}}$ = 6.39, 95% CI [6.02, 6.77], $M_{\text{increasing}}$ = 6.37, 95% CI [6.01, 6.73]), $t(258) = 0.09$, $p = .928$, $d = 0.01$, but started to diverge immediately after the first practice round. Because participants in the *decreasing-difficulty* condition saw the hard set of letters first, they reported lower self-efficacy levels (M = 4.84, 95% CI [4.45, 5.24]) than those in the *increasing-difficulty* condition (M = 6.05, 95% CI [5.70, 6.39]), $t(258) = 4.53$, $p > .001$, d = 0.56. After completing all three practice rounds, however, participants' self-efficacy was significantly higher in the *decreasing-difficulty* condition ($M = 6.56$, 95% CI [6.16, 6.96]) than in the *increasing-difficulty* condition ($M = 4.91$, 95% CI [4.50, 5.33]), $t(258) = 5.61$, $p < .001$, $d = 0.70$.

Discussion

This result directly contradicts participants' predictions about their perceived efficacy after completing word find tasks in different orders and provides the proof of concept necessary to continue investigating how completing the same tasks in different orders can change feelings of efficacy in other domains and paradigms.

Experiment 5: Analogy Task

Experiment 5 serves three purposes. As a conceptual replication of Experiment 4, I aim to show that the mispredictions about efficacy are not confined to one specific task. I also test for behavioral consequences of increased efficacy. Specifically, this experiment tests my prediction that an increase in efficacy will lead to an increase in motivation. Third, this experiment provides a test for the efficacy predictions from Experiment 2.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/u85fz).

Participants. I recruited 363 participants (159 female, $M_{\text{age}} = 35.42$ years, 95% CI [34.37, 36.47]) through Amazon Mechanical Turk who completed a survey in exchange for $0.90 with the opportunity to earn a bonus.

Design and task. Participants were randomly placed into one of three experimental conditions (between-subjects): *increasing-difficulty order, decreasing-difficulty order,* and *control*. The task in this experiment was the same analogy task used in Experiment 2. For all multiple choice analogies used in this experiment, see Appendix 3.

Manipulating task difficulty. To select the different analogies for each round, I pretested 149 analogies and chose eighteen to create three blocks of six analogies: hard ($M = 1.24$ correct, 95% CI [1.09, 1.39]), medium ($M = 2.75$ correct, 95% CI [2.54, 2.96]), and easy ($M = 5.05$ correct, 95% CI [4.84, 5.27]). See Appendix 3 for further details about the pretest. In the *decreasing-difficulty* condition, I showed participants the hard analogies, then the medium analogies, and then the easy analogies. In the *increasing-difficulty* condition, conversely, they completed the same rounds of analogies but in the opposite order—easy, then medium, then hard. In the *control condition* participants completed two easy analogies, two medium analogies, and two hard analogies in random order in each round. Importantly, all three conditions contained the same eighteen analogies, giving participants the same performance experience in aggregate. After each round, I showed participants feedback about how many analogies they answered correctly in that round and in all previous rounds.

Procedure. After participants signed the consent form, I described the task and offered an example analogy. I then told participants they would be completing three rounds of six analogies each and that they would answer a few questions after each round. I told them if they did not answer the analogy within ten seconds, it would be considered incorrect. For each analogy participants answered correctly, they earned a $0.02 bonus. After reading the instructions, participants answered four attention check questions to make sure they had read and understood the instructions (see Appendix 3 for details).

Practice analogies. Participants then saw four easy practice analogies. The first two were fill in the blank: "helicopter : ______ :: submarine : water" and "______ : eat :: tired : sleep" and I used them to exclude participants who were clearly not putting any effort into the task (pre-registered). The second two practice analogies were multiple-choice to get participants used to the format of the experiment. After the multiple choice analogies participants immediately received feedback in the form of a green check that said "YES!" (correct answer) or a big red X (incorrect answer).

Perceived efficacy. After the practice analogies and after each round, I asked three questions to assess self-efficacy: "How skilled do you think are at these analogy tasks?", "How confident do you feel about these analogy tasks?", and "How much do you trust your ability to answer these analogy tasks correctly?" (1 = *not at all [skilled/confident/]*, 10 = *very [skilled/confident/much]; α*s ≥ .95).

Memory bias. Weinstein and Roediger (2010) found that participants answering questions in an increasing-difficulty order believed they had answered more questions correctly than those answering questions in decreasing-difficulty or random orders, though actual performance did not differ. They also found that those participants remembered the tasks more optimistically and felt better about their past performance (2012). To see if this finding replicated, I asked "How many of the 18 analogies did you answer correctly? If you don't remember, please just make your best guess." and "Overall how hard did you think it was to answer those analogies correctly" (1 = *not at all difficult*, 10 = *very difficult*). I counterbalanced the order in which participants saw these two questions and the motivation questions.

Motivation. I operationalize motivation both by adapting scales from other work (e.g., Christophel, 1990) and by creating my own scales based on established guidelines (e.g., Touré-Tillery & Fishbach, 2014). The motivation-related measures in this experiment include likelihood to continue, monetary value necessary to continue, enjoyment of the task, state motivation, intrinsic motivation, and extrinsic motivation.

State motivation. To assess state motivation I asked participants to choose the number between 1 and 7 that best represented their feelings immediately following the task on nine different states of being (*Motivated–Unmotivated*, *Interested–Uninterested*, *Not stimulated–Stimulated*, *Don't want to repeat–Want to repeat*, *Inspired–Uninspired*, *Unchallenged–Challenged*, *Un-invigorated–Invigorated*, *Unenthused–Enthused*, *Excited–Not excited*; adapted Christophel, 1990).

Intrinsic vs extrinsic motivation. To separate intrinsic from extrinsic motivation, I asked participants to rate on a 7-point scale (1 = *not at all motivating*, 7 = *very motivating*) how motivating each of six factors were in their performance, the first three of which are extrinsic items and the second three of which are intrinsic items ("I wanted to earn the bonus.", "I wanted to maximize my earnings.", "I wanted to make more money.", "I wanted to prove to myself that I could do well.", "I know that I am good at these types of puzzles.", "I enjoy completing these types of puzzles."). Participants saw these six questions in random order.

Willingness to continue. I asked two questions to assess willingness to continue. First I asked "How likely would you be to repeat this task if we offered you another $0.90 to complete three more rounds of analogies (different from the ones you completed, but similar in format)? You'd get the $0.90 regardless of how many questions you get correct" (1 = *not at all likely*, 10 = *very likely*). I also asked participants "How much money in total would we need to offer in order for you to want to perform this task again (i.e., answer three more rounds of analogies similar to the ones you completed)? Imagine that you'll get the money no matter how many questions you get correct" and instructed them to "enter the amount as a simple number, e.g. 0.50 for 50 cents or 1.25 for a dollar twenty-five."

Awareness of changing task difficulty. In order to ascertain to what participants were attributing their changing performance and as a manipulation check, I asked participants to answer two slider bar questions. To determine whether they recognized the tasks were changing in difficulty, I asked "Did you think the analogies were changing in difficulty" on a slider anchored at 0 (*Yes they were getting way easier)* and 100 (*Yes they were getting way harder*) which started at 50 (*No change in difficulty).*

Awareness of changing performance. To determine whether they thought their own performance was changing, I asked "Overall, were you getting better at analogies, getting worse, or staying about the same?" and participants moved a slider anchored at 0 (*Getting much worse*) and 100 (*Getting better*) which started at 50 (*Staying the same*).

Control variables. To control for participants' experience with the task, I asked about task familiarity: "How familiar are you with analogy tasks similar to the ones you completed today?" (*I have never played a game like that before*, *I have played a game like that a few times*, *I sometimes play games like that*, *I frequently play games like that*, *I play games like that almost every day*). To control for aversiveness, I also asked "Overall, how much do you enjoy engaging in these analogy tasks?" (1 = *not at all*, 10 = *very much*). Finally I collected education, income, employment, age, and gender.

Results

Analysis strategy. I pre-registered that I would remove any participants who clearly did not try to complete the two fill in the blank practice analogies correctly and participants who answered both of the extremely easy practice multiple choice analogies incorrectly. After these exclusions, 327 participants remained (147 female, M_{age} = 35.63 years, 95% CI [34.51, 36.76]) of the 363 who completed the survey, though the results do not differ when I include those participants. After exclusions, the *increasing-difficulty* condition includes 108 participants for analysis, the *decreasing-difficulty* condition has 112 and the *control* condition has 107 remaining participants.

Practice round scores. In order to rule out an alternate explanation that this manipulation actually improves skill in one condition as compared to the others, which would account for any differences in ratings of efficacy, I compared the aggregate practice round scores. These scores were no different in the *increasing-difficulty* (M = 10.23, 95% CI [9.86, 10.61]), *decreasing-difficulty* (M = 10.60, 95% CI [10.22, 10.97]) or *control* conditions (M = 10.39, 95% CI [9.99, 10.79]), $F(2, 324) = 0.92$, $p = .400$, $\eta^2 < 0.01$, suggesting that my manipulation did not affect actual skill levels. As designed, and replicating my pretest results, participants correctly answered more analogies in the easy round (M = 5.64, 95% CI [5.54, 5.75]) than the medium round (M = 3.56, 95% CI [3.38, 3.75]), $t(438) = 19.26$, $p < .001$, $d = 1.84$, and more analogies in the medium round than in the hard round (M = 1.21, 95% CI [1.08, 1.35]), $t(438) = 20.37$, $p < .001$, $d = 1.95$.

Perceived efficacy. As in Experiment 4, condition significantly impacted perceived efficacy after the third practice round, $F(2, 324) = 12.66$, $p < .001$, $\eta^2 = 0.07$, with participants in the *decreasing-difficulty* condition (M = 6.13, 95% CI [5.70, 6.56]) reporting feeling significantly more efficacious than those in the *increasing-difficulty* condition (M = 4.62, 95% CI [4.15, 5.08]), $t(218) = 4.75$, $p < .001$, $d = 0.64$. Participants' efficacy in the *control* condition (M = 5.26, 95% CI [4.89, 5.64]) fell in between those in the *decreasing-difficulty* condition, $t(217) = 2.99$, $p = .003$, $d = 0.41$, and those in the *increasing-difficulty* condition, $t(213) = 2.15$, $p = .033$, $d = 0.29$.

Memory bias. There was no difference by condition on how accurately participants remembered how many analogies they had answered correctly, $F(2, 324) = 2.10$, $p = .125$, $\eta^2 = 0.01$. Perception of overall difficulty did differ by condition though, $F(2, 324) = 4.85$, $p = .008$, $\eta^2 = 0.03$, in that participants in the *decreasing-difficulty* condition ($M = 6.50$, 95% CI [6.19, 6.81]) found the task easier in retrospect than those in the *increasing-difficulty* ($M = 7.06$, 95% CI [6.73, 7.38]) or *control* conditions ($M = 7.09$, 95% CI [6.83, 7.35]), ts > 2.44, ps < .016, ds > 0.33, though this could simply be a recency effect since those in the *decreasing-difficulty condition* saw the easiest analogies immediately before answering that question.

State motivation. There was no difference by condition in state motivation, $F(2, 324) = 0.73$, $p = .483$, $\eta^2 < 0.01$.

Intrinsic vs extrinsic motivation. There was no difference by condition in either intrinsic, $F(2, 324) = 0.35$, $p = .704$, $\eta^2 < 0.01$, or extrinsic motivation, $F(2, 324) = 1.87$, $p = .155$, $\eta^2 = 0.01$, though generally participants report being more extrinsically motivated ($M = 5.30$, 95% CI [5.11, 5.48]) than intrinsically motivated ($M = 4.88$, 95% CI [4.73, 5.04]) on this task, $t(652) = 3.40$, $p < .001$, $d = 0.27$, so it is possible that the increased efficacy did not affect motivation because extrinsic motivation drove this specific task.

Willingness to continue. There was no difference by condition for how likely participants would be to complete the task again for the same price, $F(2, 324) = 0.90$, $p = .407$, $\eta^2 < 0.01$. Interestingly, the overall mean was above nine on a ten point scale ($M = 9.24$, 95% CI [9.04, 9.43]), indicating that nearly all of the participants would be happy to participate in this survey again for the same price. There was a difference by condition in how much participants would need to be paid to complete the task again, $F(2, 324) = 4.52$, $p = .012$, $\eta^2 = 0.03$, in that those in the *decreasing-difficulty* condition ($M = 1.19$, 95% CI [1.00, 1.38]) would require more money to retake this survey than those in the *increasing-difficulty* ($M = 0.97$, 95% CI [0.88, 1.06]) or *control* conditions ($M = 0.93$, 95% CI [0.86, 1.00]), ts > 2.03, ps < .043, ds > 0.28.

Correlational analyses. There were significant correlations between efficacy and both state motivation, $r(325) = 0.26$, $p < .001$, and intrinsic motivation, $r(325) = 0.24$, $p < .001$. Those motivation values also correlated highly with one another, $r(325) = 0.64$, $p < .001$, and with willingness to continue, rs(325) > 0.17, ps < .002. Extrinsic motivation was significantly correlated with the other three motivation variables, rs(325) > 0.18, ps < .001, but neither extrinsic motivation, $r(325) = -0.02$, $p = .660$, nor willingness to continue, $r(325) = -0.01$, $p = .877$, was significantly correlated with efficacy. For a complete correlation plot of these five variables, see Appendix 3.

Awareness of changing task difficulty. Those in the *increasing-difficulty* condition recognized that the tasks were getting harder ($M = 93.23$, 95% CI [90.61, 95.85]), whereas those in the *decreasing-difficulty* condition recognized that the tasks were getting easier ($M = 20.71$, 95% CI [15.69, 25.72]), and the two conditions reported significantly different assessments of task-difficulty, $t(218) = 25.14$, $p < .001$, $d = 3.41$. In the *control* condition, participants thought the tasks were getting slightly harder ($M = 64.48$, 95% CI [60.93, 68.02]), but not nearly as hard as in the *decreasing-difficulty* condition, $t(213) = 12.95$, $p < .001$, $d = 1.77$. Thus, participants were aware that the task difficulty was changing. This also serves as a manipulation check.

Awareness of changing performance. Participants also attributed their changing scores internally, despite their awareness that the tasks were changing in difficulty. They reported getting better at analogies in the *decreasing-difficulty* condition ($M = 75.28$, 95% CI [72.04, 78.51]) and getting worse at analogies in the *increasing-difficulty* condition ($M = 28.76$, 95% CI [23.78, 33.74]), $t(218) = 15.64$, $p < .001$, $d = 2.12$. In the control condition, participants did not think their skill was changing ($M = 50.08$, 95% CI [46.54, 53.63]), *one-sided* $t(106) = 0.05$, $p = .963$, $d = 0.01$.

Discussion

This experiment replicated the finding from Experiment 5 that completing tasks in decreasing-difficulty order creates more reported efficacy than completing them in increasing-difficulty order and contradicts the predictions from Part 1, specifically from Experiment 2. Contradicting one of my hypotheses, however, and despite the difference in felt efficacy, participants reported no difference in motivation, whether they completed the tasks in increasing-difficulty, decreasing-difficulty, or random order, though there were correlations between efficacy and motivation. However, because task order does not seem to causally impact motivation even when it changes efficacy as predicted, for most of the rest of this chapter I will focus on mispredictions and mechanisms instead of the downstream effects of increased efficacy.

Experiment 6: Bouncing Ball Task

This experiment examines whether task ordering has an impact on perceived efficacy outside the realm of word-related tasks. In it, I ask participants to predict the path of a bouncing ball and determine whether that path will lead it through a goal. I chose this task to demonstrate the robustness of this effect outside of verbal-reasoning tasks and tasks that benefit from previous knowledge such as finding words and completing analogies.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/dfqut).

Participants. I recruited 603 participants (294 female, $M_{\text{age}} = 36.51$ years, 95% CI [35.58, 37.44]) through Amazon Mechanical Turk who completed a survey in exchange for \$1.00 with the opportunity to earn a bonus.

Design and task. In this experiment, participants played a bouncing ball game, in which they had to guess whether a ball's trajectory would take it through a goal based on a small initial trajectory. To complete the task, participants had to extrapolate the trajectory of the ball using the placement of the ball and a small guide line indicating its initial movement. For the instruction graphic and an example of an easy trial and a hard trial, see Appendix 3. I used a 3 (task order: increasing-difficulty, decreasing-difficulty, or random) × 2 (time: 1 second or 10 seconds) between-participants design. All participants saw 21 trials broken up into three rounds but in different orders depending on their randomly selected task order condition, mimicking the paradigm used in Experiments 4 and 5. Participants were also randomly selected into a time condition and had the opportunity to score an extra half point per trial by answering their trials correctly within their specified time limit. I originally ran this experiment as part of a research question about momentum, and I thought perhaps time pressure would increase perceived momentum. This manipulation is not relevant to the current research and does not impact efficacy ratings, so I will not discuss it further.[13]

Procedure. After signing a consent form, I told participants "In this experiment, you will see a ball bouncing around in a box. There are four walls on the outside and one in the middle. The wall in the middle has a hole in it. Your task is to determine whether the ball will go through the hole *before it hits the center wall.*" I then showed them a picture of the task (see Appendix 3) and told them that they would be making the prediction about whether the ball would go through

[13] The time condition has no impact on efficacy ratings after the third practice round for all participants ($M_{1\text{sec}} = 3.97$, 95% CI [3.80, 4.13], $M_{10\text{sec}} = 4.09$, 95% CI [3.94, 4.24]), $t(601) = 1.08$, $p = .281$, $d = 0.09$, or more importantly, for participants in the *increasing-difficulty* and the *decreasing-difficulty* conditions ($M_{1\text{sec}} = 4.01$, 95% CI [3.81, 4.22], $M_{10\text{sec}} = 3.95$, 95% CI [3.76, 4.13]), $t(403) = 0.49$, $p = .623$, $d = 0.05$.

the hole in a number of different trials. I also told participants that they would earn points for accuracy and speed. Participants then answered a few attention check questions (see Appendix 3) before moving into three instruction trials. After the instruction trials, participants answered a few questions before moving into the three practice rounds. After each trial in the practice rounds, I told participants whether they answered correctly or incorrectly and showed them the full path of the ball. After each practice round I told them their score in that round and reminded them of their score(s) in the previous round(s) and then they answered more questions. After all three practice rounds, I told participants that they would complete a competition round with another seven trials that were about as difficult as the average trial they had completed during the practice rounds. I incentivized performance by offering a $0.04 bonus for each point, up to a total bonus of $0.42. I then told participants that I selected a target score for them based on their performance up to this point and said if they beat their target score they earned an extra $0.05. The target score was the average number of points they had earned in each of their first three rounds. After the competition I reported to participants their scores and bonus earnings and then collected some final measures before the end of the survey.

Manipulating task difficulty. I pretested 52 stimuli and selected 31 of them for this experiment: 3 in the instruction round, 7 in each of three practice rounds, and 7 in a final competition round. For the easy round, I selected trials with an expected value of 5.75 out of 7 correct. For the medium round, I selected trials with an expected value of 4.34 out of 7, and for the hard round I selected trials with an expected value of 2.15 out of 7. I wanted the competition round to be medium difficulty, so I selected trials that had an expected value of 4.17 out of 7 correct. Once I selected the seven trials for each round, I arranged them within their selected round in a random order and then combined the rounds such that I had an ordered list of all 21 trials for the practice rounds. In the *increasing-difficulty* condition I showed participants trials 1–21 broken into three blocks of seven. In the *decreasing-difficulty* condition I showed participants trials 21–1 broken into three blocks of seven. In the *control* condition I showed participants the 21 trials in random order broken into three blocks of seven. For more details about the completion rates of each trial during the pretest, see Appendix 3.

Feedback. After each round I told participants their total number of points (out of a possible 10.5) and broke down that value into how many trials they answered correctly (out of a possible 7) and how many extra points they earned for answering within their time limit (out of a possible 3.5). I also reminded them of their score(s) from each previous round.

Perceived efficacy. After the instruction round and after each of the practice rounds I asked “How skilled do you think you are at this task?” (1 = *not at all skilled*, 7 = *very skilled*). Based on the alphas from previous experiments in this work, I determined that asking only about skill appropriately approximates perceived efficacy. Because this survey was already quite long, I chose to ask this one question instead of the three-question index I usually collect.

Perceived momentum. To measure perceived momentum, I created a three-question momentum index and asked all three questions after each of the three practice rounds: “After completing this round, how much do you feel like you are gaining or losing momentum in the task, from -50 (lost momentum) to 50 (gained momentum)?”, “After completing this round, how much do you feel like you are “gathering or losing steam”, from -50 (losing a lot of steam) to 50 (gathering a lot of steam)?”, and “After completing this round, how much do you feel like you are “on a roll”, from -50 (not at all on a roll) to 50 (very much on a roll)?” Participants answered these questions by moving a slider bar that was initially placed at zero. I collapsed these questions into one momentum index for analysis (αs > 0.95).

Enjoyment. After each practice round, I asked "How much are you enjoying this game right now?" (1 = *not enjoying it at all*, 7 = *enjoying it a lot*).

Changing performance. After the second and third practice round I asked participants "How do you think your performance is changing over time?" (-3 = *I'm getting a lot worse*, 0 = *No change*, 3 = *I'm getting a lot better*).

Expected performance measures. After I described the competition, I asked participants for a number of predictions about their performance. I asked both "How likely are you to beat your target score?" (1 = *not at all likely*, 10 = *extremely likely*) and "Do you think you will beat your target score?" (*Yes*, *No*). I also asked for specific performance predictions. I asked "How many points do you think you will get in the competition?" (0–10.5) and "How many trials do you think you will get correct in the competition?" (0–7). To incentivize accuracy, I offered a bonus of $0.02 if participants' final score was within two points of their predicted score and another $0.02 if the number of trials they answered correctly was within one trial of their prediction.

Control variables and manipulation check. After the competition, participants finished the survey by answering a few control variables and demographic questions. I asked "How much fun did you think that game was?" (1 = *not at all fun*, 7 = *extremely fun*), "In Round 1[/2/3], how difficult did you find it to earn all the points?" (1 = *not at all difficult*, 7 = *extremely difficult*), and "More generally, did it seem to be getting easier or harder to earn points throughout the three practice rounds (before the competition)?" (1 = *getting quite a bit easier*, 4 = *neither easier nor harder*, 7 = *getting quite a bit harder*). I also asked "How much experience do you have playing games of this sort?" (*I never play games like this*, *I rarely play games like this*, *I sometimes play games like this*, *I often play games like this*, *I always play games like this*), age, and gender.[14]

Results

Analysis plan. Since my current interest in this data is only in ratings of efficacy, and because the time condition had no effect on efficacy ratings, I am going aggregate the data into three conditions for analysis: *increasing-difficulty*, *decreasing-difficulty*, and *control*.

Practice round scores. Participants gained slightly fewer points and answered fewer trials correctly in the *decreasing-difficulty* condition (M_{points} = 5.73, 95% CI [5.56, 5.90], M_{trials} = 12.33, 95% CI [12.01, 12.64]) than in the *control* condition (M_{points} = 5.98, 95% CI [5.80, 6.16], M_{trials} = 12.96, 95% CI [12.66, 13.27]) or the *increasing-difficulty* condition (M_{points} = 6.02, 95% CI [5.86, 6.18], M_{trials} = 12.89, 95% CI [12.58, 13.19]), ts > 1.99, ps < .047, ds > 0.20. If anything, this offers a conservative test of my hypothesis that participants completing tasks in decreasing-difficulty order will report more efficacy after three practice rounds than those completing the tasks in increasing-difficulty order, since participants in the *decreasing-difficulty* condition performed objectively worse over those three rounds than participants in the *increasing-difficulty* condition.

Perceived efficacy. As in Experiments 4 and 5, condition significantly impacted perceived efficacy after the third practice round, $F(2, 600) = 24.26$, $p < .001$, $\eta^2 = 0.07$, with participants in the *decreasing-difficulty* condition (M = 4.45, 95% CI [4.25, 4.64]) reporting feeling significantly more efficacious than those in the *increasing-difficulty* condition (M = 3.53, 95% CI [3.35, 3.70]), $t(403) = 6.89$, $p < .001$, $d = 0.69$. Participants' efficacy in the *control* condition (M = 4.13, 95% CI [3.94, 4.32]) fell in between those in the *decreasing-difficulty*

[14] Prior to completing the practice rounds, participants also completed a growth mindset scale (eight items, modified to pertain to performance; De Castella & Byrne, 2015) which was included to test a theory unrelated to the present work.

condition, $t(396) = 2.30$, $p = .022$, $d = 0.23$, and those in the *increasing-difficulty* condition, $t(401) = 4.56$, $p < .001$, $d = 0.46$. Interestingly and unexpectedly, every single control variable (experience with the task, how fun they found the exercise, how much they were enjoying the task after the third round, age, and gender) also predicted ratings of efficacy after the third practice round. However, when I include all of these variables with condition in a multiple regression model using only the *increasing-difficulty* and *decreasing-difficulty* conditions, condition still significantly predicts efficacy ratings ($\beta = 0.63$, $t = 5.39$, $p < .001$), $F(6, 397) = 41.68$, $p < .001$, $R^2 = 0.39$, indicating that the order of tasks causally influences efficacy even when controlling for enjoyment, fun, experience, age, and gender.

Awareness of changing task difficulty. Participants indicated a significant difference by condition in whether it seemed to be getting easier or harder to earn points throughout the three practice rounds, $F(2, 600) = 153.8$, $p < .001$, $\eta^2 < 0.34$. Those in the *increasing-difficulty* condition ($M = 5.39$, 95% CI [5.22, 5.55]) thought the tasks were getting more difficult over time than those in the *control* condition ($M = 4.13$, 95% CI [3.93, 4.32]), $t(401) = 9.60$, $p < .001$, $d = 0.96$, while those in the *decreasing-difficulty* condition ($M = 3.05$, 95% CI [2.86, 3.24]) thought the tasks were getting easier over time as compared to the *control* condition, $t(396) = 7.67$, $p < .001$, $d = 0.77$, indicating that our manipulation was successful.

Awareness of changing performance. Participants also attributed their changing scores internally, despite their awareness that the tasks were changing in difficulty. After the third practice round, participants in the *increasing-difficulty* condition ($M = -1.46$, 95% CI [-1.64, -1.28]) thought they were getting worse at the bouncing ball task, *one-sided* $t(204) = 16.16$, $p < .001$, $d = 2.26$. Those in the *control* condition did not think their performance was changing ($M = 0.10$, 95% CI [-0.10, 0.30]), *one-sided* $t(197) = 0.99$, $p = .323$, $d = 0.14$, and participants in the *decreasing-difficulty* condition thought they were getting better at the task ($M = 1.57$, 95% CI [1.40, 1.73]), $t(199) = 18.82$, $p < .001$, $d = 2.67$.

Remaining variables. Since I collected the remaining dependent variables (growth mindset, perceived momentum, and performance expectations) in order to test an entirely different hypothesis, they are not relevant to this work and I will not report them here. If you are interested in those results, please contact the author.

Discussion

Experiment 6 provides further support for the hypothesis that completing tasks in decreasing-difficulty order creates a greater sense of efficacy than completing them in increasing-difficulty order. This experiment shared a paradigm with both the word find of Experiment 4 and the analogies of Experiment 5 in that participants completed three rounds of approximate difficulty level of easy, medium, and hard, and answered questions about their efficacy after each round. This experiment goes further, however, in demonstrating the effect in a task that did not require any prior knowledge and moved away from the domain of verbal reasoning. In the next two experiments, I will remove the rounds from the paradigm and simply show a number of tasks in increasing-difficulty order or decreasing-difficulty order. In the new paradigm, participants will only report their efficacy feelings once, after they have completed all of the tasks. I changed the paradigm to see if this effect generalizes away from rounds, and I chose different tasks in hopes of continuing to improve the external validity of these findings.

Experiment 7: Non-verbal Reasoning Task

For Experiment 7, I chose a non-verbal reasoning task and attempted to replicate my previous findings in a different paradigm. Participants saw ten total tasks in different orders and then reported how they felt after completing them. In this experiment, I also included the motivation questions from Experiment 5 for thoroughness.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/9syf4).

Participants. I recruited 299 participants (117 female, 4 gender non-binary, $M_{\text{age}} = 35.31$ years, 95% CI [34.15, 36.47]) through Amazon Mechanical Turk who completed a survey in exchange for $0.80 with the opportunity to earn a bonus.

Design and task. Participants were randomly placed into one of three experimental conditions (between-subjects): *increasing-difficulty order*, *decreasing-difficulty order*, and *control*. The task in this experiment was to complete ten non-verbal reasoning questions. For each question, I showed participants four squares with various markings in them and asked them to choose from five options a square which completes the pattern. See Figure 1 for an example.

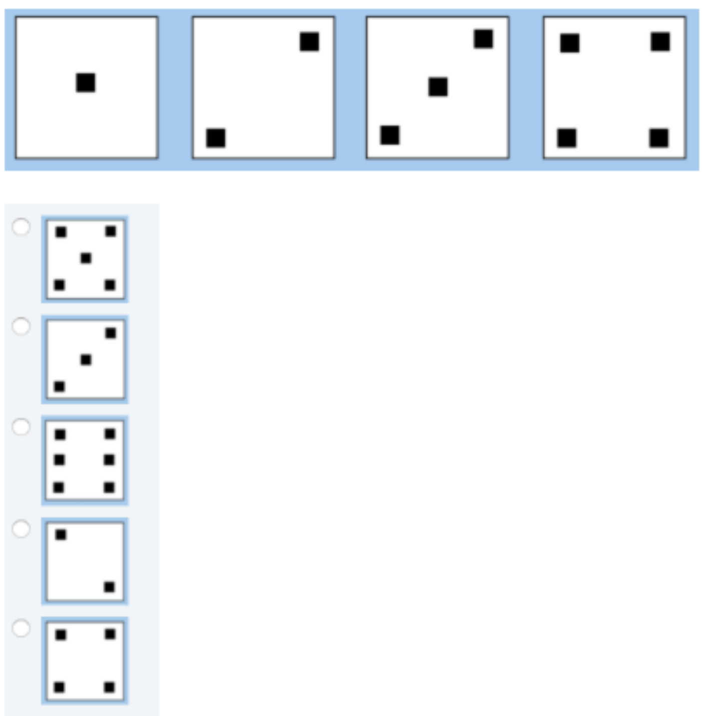


Figure 1. Example of non-verbal reasoning question

Procedure. After signing the consent form, I described the task to participants and offered the example question above. I told them "You would want to pick **the top option**, because **it has five dots**. That completes the pattern. The first square has one dot, the second square has two dots, the third square has three dots, the fourth square has four dots, so the fifth square should have five dots." I then told them that was easier than most of the questions they would see in the experiment. I also showed them an example of a hard question and explained the correct answer. I then told participants they would see ten non-verbal reasoning questions and have as long to answer each one as they'd like. After each question, I immediately provided feedback about whether the answer was correct or incorrect. I also incentivized effort by offering a bonus of $0.03 for each correct answer. I then asked participants some attention check questions and some efficacy questions. After they completed the ten trials, they answered a number of other questions before finishing the experiment.

Manipulating task difficulty. To order the questions by difficulty, I ran a pretest. I chose ten questions from the pretest for the experiment. Of those ten questions, 86% of participants answered the easiest question correctly while only 24% of participants answered the most difficult question correctly. See Appendix 3 for further details about the pretest and the stimuli used. In the *decreasing-difficulty* condition, I showed participants the most difficult question first and each subsequent question was easier than the last. In the *increasing-difficulty* condition, conversely, participants saw the ten questions in the opposite order; they saw the easiest question first and each subsequent question was more difficult than the last. In the *control condition* participants saw the ten trials in a random order. Importantly, all three conditions contained the same ten non-verbal reasoning questions, giving participants the same set of experiences in aggregate.

Perceived efficacy. After I described the task and after they completed the ten questions, I asked three questions to assess self-efficacy: "How skilled do you think are at solving these non-verbal reasoning tasks?", "How confident do you feel about solving these non-verbal reasoning tasks?", "How much do you trust your ability to solve these non-verbal reasoning tasks?" (1 = *not at all [skilled/confident/],* 10 = *very [skilled/confident/much]*; αs $\geq .97$).

Memory bias. To see if memory bias differed by condition, I asked "How many of the 10 questions did you answer correctly? If you don't remember, please just make your best guess" and "Overall how hard did you think those ten questions were to answer correctly" (1 = *not at all difficult*, 10 = *very difficult*). I counterbalanced the order in which participants saw these two questions and the motivation questions.

Motivation. I used the same questions as in Experiment 5 to determine participants' motivation and willingness to continue after completing the task.

Awareness of changing task difficulty. In order to ascertain to what participants were attributing their changing performance and as a manipulation check, I asked participants to answer two slider bar questions. To determine whether they recognized the tasks were changing in difficulty, I asked "Did you think the non-verbal reasoning questions were changing in difficulty" on a slider anchored at 0 (*Yes they were getting way easier*) and 100 (*Yes they were getting way harder*) which started at 50 (*No change in difficulty*).

Awareness of changing performance. To determine whether they thought their own performance was changing, I asked "Overall, were you getting better at non-verbal reasoning over time, getting worse, or staying about the same?" and participants moved a slider anchored at 0 (*Getting much worse*) and 100 (*Getting better*) which started at 50 (*Staying the same*).

Control variables. To control for participants' experience with the task, I asked "How familiar are you with non-verbal reasoning tasks similar to the ones you completed today?" (*I have never played a game like that before*, *I have played a game like that a few times*, *I sometimes play games like that*, *I frequently play games like that*, *I play games like that almost every day*). To control for aversiveness, I also asked "Overall, how much do you enjoy engaging in these non-verbal reasoning tasks?" (1 = *not at all*, 10 = *very much*). Finally I collected education, income, employment, age, and gender.

Results

Analysis strategy. I pre-registered that I would remove any participants who spent less than an average of one second on each question. Only six participants were excluded by this metric, leaving 293 participants (114 female, 4 non-binary, M_{age} = 35.513 years, 95% CI [34.33, 36.68]).

Overall score. Aggregate scores were no different in the *increasing-difficulty* (M = 5.90, 95% CI [5.50, 6.30]), *decreasing-difficulty* (M = 5.51, 95% CI [4.98, 6.03]) or *control* conditions (M = 5.58, 95% CI [5.15, 6.01]), $F(2, 290) = 0.82$, $p = .440$, $\eta^2 < 0.01$, suggesting that the manipulation did not affect actual skill levels. As designed, and replicating the pretest results, the questions changed in difficulty linearly in the *increasing-difficulty* and *decreasing-difficulty* conditions.

Perceived efficacy. Unlike in previous experiments, efficacy did not vary by condition, $F(2, 290) = 0.22$, $p = .801$, $\eta^2 < 0.01$. After completing all ten questions, participants in the *decreasing-difficulty* condition (M = 4.90, 95% CI [4.40, 5.40]), *increasing-difficulty* condition (M = 5.13, 95% CI [4.66, 5.59]), and *control* condition (M = 5.00, 95% CI [4.54, 5.46]) reported similar levels of efficacy.

Memory bias. There was also no difference by condition on how accurately participants remembered how many analogies they had answered correctly, $F(2, 290) = 1.47$, $p = .232$, $\eta^2 = 0.01$, or perception of overall difficulty, $F(2, 290) = .98$, $p = .376$, $\eta^2 < 0.01$.

State motivation. There was no difference by condition in state motivation, $F(2, 290) = 1.01$, $p = .467$, $\eta^2 < 0.01$.

Intrinsic vs extrinsic motivation. There was no difference by condition in either intrinsic, $F(2, 290) = 0.99$, $p = .372$, $\eta^2 < 0.01$, or extrinsic motivation, $F(2, 290) = 0.68$, $p = .508$, $\eta^2 < 0.01$, and as in the analogy task, participants report being more extrinsically motivated (M = 4.84, 95% CI [4.62, 5.05]) than intrinsically motivated (M = 4.28, 95% CI [4.10, 4.46]) on this task, $t(584) = 3.93$, $p < .001$, $d = 0.33$.

Willingness to continue. There was no difference by condition for how likely participants were to complete the task again for the same price, $F(2, 290) = 1.67$, $p = .190$, $\eta^2 = 0.01$. As in the analogy task, the overall mean was almost nine on a ten point scale (M = 8.96, 95% CI [8.73, 9.19]), indicating that nearly all of the participants would be happy to participate in this survey again for the same price. There was also no difference by condition in how much participants would need to be paid to complete the task again, $F(2, 290) = 1.00$, $p = .371$, $\eta^2 < 0.03$.

Correlational analyses. In this experiment, all four motivation variables and ratings of efficacy were significantly correlated with one another, rs(291) > 0.13, ps > .028. For a complete correlation plot of these five variables, see Appendix 3.

Awareness of changing task difficulty. Those in the *increasing-difficulty* condition recognized that the tasks were getting harder (M = 80.56, 95% CI [77.74, 83.38]), whereas those in the *decreasing-difficulty* condition recognized that the tasks were getting easier (M = 40.96, 95% CI [35.61, 46.31]), and the two conditions reported significantly different assessments of task-difficulty, $t(193) = 13.04$, $p < .001$, $d = 1.88$. In the *control* condition (M = 64.12, 95% CI [59.41, 68.44]), participants thought the tasks were getting slightly harder, but not nearly as hard as in the *decreasing-difficulty* condition, $t(194) = 6.33$, $p < .001$, $d = 0.91$. Thus, participants were aware that the task difficulty was changing. This also serves as a manipulation check.

Awareness of changing performance. Participants also attributed their changing scores internally, despite their awareness that the tasks were changing in difficulty. They reported getting better at non-verbal reasoning questions in the *decreasing-difficulty* condition (M = 62.28, 95% CI [58.33, 66.25]) and getting slightly worse at non-verbal reasoning questions in the *increasing-difficulty* condition (M = 46.15, 95% CI [42.11, 50.19]), $t(193) = 5.66$, $p < .001$, $d = 0.82$. In the *control* condition, participants did not think their skill was changing (M = 53.37, 95% CI [49.38, 57.76]), *one-sided* $t(97) = 1.69$, $p = .094$, $d = 0.34$.

Discussion

In Experiment 7, task order did not have any impact on reported efficacy, suggesting a potential boundary condition for the ordering effect in Experiments 4–6. It is possible that asking participants to consider their efficacy repeatedly is integral for task ordering to change felt efficacy because it forces them to attend to the trajectory. It is also interesting to note that while completing tasks in decreasing-difficulty order did not create more efficacy than completing them in increasing-difficulty order as it did in the first three experiments in Part 2, the opposite was also not true, which offers further evidence of a misprediction in Part 1.

Experiment 8: Job Application Task

For Experiment 8, I used a similar paradigm to Experiment 7 but a completely new task. I asked participants to complete five common exercises required to apply for a job. I chose this task to establish external validity by demonstrating the effect in a more life-relevant domain than online games. This experiment also serves as further investigation into whether reporting efficacy multiple times is a potential boundary condition of the effect of task ordering on efficacy.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/am36v).

Participants. I recruited 200 third and fourth year students (150 female, 2 non-binary, M_{age} = 20.96 years, 95% CI [20.78, 21.13]) from the University of California, Berkeley's Experimental Social Science Laboratory (XLab) through SONA. I chose third and fourth year students to maximize the chance that my participants had some experience applying to jobs.

Design and task. Participants were randomly placed into one of two experimental conditions (between-subjects): *increasing-difficulty order* and *decreasing-difficulty order*. The task in this experiment was to complete five pre-determined tasks related to applying for a job. I selected five tasks from a pretest of nine options. The tasks were: indicating times the participant would be available for a phone interview, filling out a form with information about three previous jobs, detailing three previous job duties, creating an objective sentence for a resume, and writing about a challenge participants had faced in the past. For more details on the pretest and the tasks, see Appendix 3. I chose tasks that were rated as largely the same on pleasantness, importance, and desire to complete them, but varied in ratings of difficulty. If there is an effect of task ordering on efficacy, these pretest ratings should hopefully help rule out those confounding constructs as alternative explanations for the effect.

Procedure. Participants came into the lab and were given a link to a survey on Qualtrics. I told participants that they would be completing the survey while I or some research assistants wandered around the room and watched over their shoulders. I brought them into the lab so that I could make sure they stayed on task. Participants first saw all five tasks in random order and rated them on difficulty, pleasantness, importance and desire to complete them. After rating all the tasks, participants ranked them from easiest to hardest and indicated their order preference. I then randomly selected participants into either the *increasing-difficulty* or the *decreasing-difficulty* condition and told them which condition they were in. They proceeded to complete the tasks in their assigned order. After completing each task, they rated it again on difficulty, pleasantness, and importance. After completing all five tasks, they answered a few more questions and were asked to wait in the lab until all other participants had finished.

Manipulating task difficulty. Because people have different internal ratings of difficulty, instead of assigning a difficulty value to each task and displaying them according to that order, I let participants determine their own rankings to better align with the realities of our lives. Once participants ranked the tasks in order of difficulty, I randomly assigned them to either the *increasing-difficulty* condition or the *decreasing-difficulty* condition and then they completed the five tasks in their prescribed order, based on their own rankings.

Task ratings. Prior to completing the tasks, participants rated them on the following metrics: “How difficult this task is”, “How pleasant this task is”, “How important this task is in the job application process”, “How much you want to complete this task” (1 = *not at all*, 10 = *very*). After completing the tasks, participants answered the first three questions again but did not answer how much they wanted to complete the task since they already had completed it.

Task rankings. I told participants “Next you will actually complete the tasks you just rated. The tasks below are shown in random order. Please rank order them from "easiest" (1) to "hardest" (5), using whatever metrics you would like to make that determination. You can move tasks by clicking on them and then dragging to re-position within the rankings.”

Preferences. I also asked participants to indicate their preference for task ordering—“You will be completing these tasks either from the easiest to the hardest or from the hardest to the easiest, as you just ranked them. If you had a choice, would you prefer to start with the easiest and end with the hardest (increasing in difficulty), or start with the hardest and end with the easiest (decreasing in difficulty)?” (*Easy to hard (increasing in difficulty)*, *Hard to easy (decreasing in difficulty)*)—and to report why they made that choice.

Perceived efficacy. As in the previous experiments in Part 2, I asked three questions to assess self-efficacy. In this experiment, though, I asked two different sets of efficacy questions, one about their perceived efficacy at applying for jobs and one about their perceived efficacy at receiving an interview. Since applying to jobs involves a number of different skills and competencies, I wanted to collect measures regarding both efficacy about the process and efficacy about the outcome. The questions were “After completing these tasks, how skilled do you think you are *at applying for jobs [getting an interview]*?”; “After completing these tasks, how confident do you feel *about applying for jobs [getting an interview]*?”; “After completing these tasks, how much do you trust your ability *to apply for jobs [get an interview]*?” (1 = *not at all [skilled/confident/]*, 10 = *very [skilled/confident/much]*; αs > .95).

Changing task difficulty. I asked participants “Did you think the tasks were changing in difficulty?” To answer, they moved a slider anchored at 0 (*Yes they were getting way easier*) and 100 (*Yes they were getting way harder*) which started at 50 (*No change in difficulty*). Since I explicitly told participants in which order they would see the tasks, this mostly serves as a manipulation check.

Other dependent variables. I collected a number of other dependent variables that I thought might differ by condition or be affected by changing efficacy. I asked “How easy do you think it would be to apply for a job right now?” (1 = *very difficult*, 10 = *very easy*), “How soon would you be willing to apply for a job?” (*Immediately*, *Tomorrow*, *Next week*, *Next month*, *At least two months from now*), “How likely is it that your next job application will be successful?” (1 = *not at all likely*, 10 = *very likely*), and “Compared to the best version of yourself, how well do you think this application package represents you?” (1 = *not at all well*, 10 = *perfect representation*).

Performance. Two blind coders rated four of the tasks on performance using a 1–10 scale. They did not rate the scheduling question since there is no way to determine how accurately participants portrayed their own schedules. I asked the coders to imagine they were reading these responses on a job application and rate the overall quality of each answer. They made their rating based on both how well participants answered the prompt and how much effort they seem to have put into their response (1 = *extremely low quality*, 10 = *extremely high quality*).

Control variables. I collected a number of control variables in this experiment. I asked "How personally useful did you find this exercise?" (1 = *not at all useful*, 10 = *extremely useful*), "How distracted were you when completing this exercise" (1 = *not at all distracted*, 10 = *extremely distracted*), and "How much effort did you put into completing these tasks?" (1 = *no effort at all*, 10 = *maximum effort*). I also collected two measure of experiences: "How familiar are you with applying for jobs?" (*I have never applied to a job before*, *I have applied to a few jobs*, *I have applied to many jobs*) and "Are you currently in the process of applying for jobs or internships or do you plan to apply for jobs or internships within the next couple of months?" (*Yes*, *No*). Finally I collected year in school, major, age, and gender.

Results

Ratings of task difficulty. I took an average of all the individual difficulty measures to create a metric of overall task difficulty. After participants completed the tasks, overall task difficulty did not differ by condition ($M_{\text{decreasing}} = 3.89$, 95% CI [3.64, 4.13], $M_{\text{increasing}} = 3.99$, 95% CI [3.71, 4.26]), $t(198) = 0.52$, $p = .602$, $d = 0.07$, indicating that participants had similar experiences in regard to difficulty.

The only specific task that differed by condition was describing a challenge or mistake they had made in a previous job, which was consistently rated the most difficult task. Those in the *increasing-difficulty* condition ($M = 5.85$, 95% CI [5.40, 6.29]) actually found the challenge to be more difficult than those in the *decreasing-difficulty* condition ($M = 5.08$, 95% CI [4.66, 5.50]), $t(198) = 2.48$, $p = .014$, $d = 0.35$, which provides a new argument for starting with your most difficult task; it may end up feeling less difficult overall. It is also interesting to note that even the most difficult task participants completed was only rated about five on a ten point scale. It is possible these tasks did not vary enough in difficulty to see the effect on efficacy that we have seen in previous experiments. For a thorough reporting of task ratings from both before and after task completion, please see Appendix 3.

Order preference. As in Experiments 1–3, participants overwhelmingly prefer to order the tasks in increasing-difficulty order with 65.5% of participants selecting easiest to hardest and the remaining 34.5% indicating a preference for decreasing-difficulty order, $t(398) = 6.50$, $p < .001$, $d = 0.65$.

Perceived efficacy. Participants' reported self-efficacy did not differ by condition regarding either applying for jobs ($M_{\text{decreasing}} = 6.23$, 95% CI [5.84, 6.62], $M_{\text{increasing}} = 6.04$, 95% CI [5.64, 6.43]), $t(198) = 0.67$, $p = .506$, $d = 0.09$, or getting an interview ($M_{\text{decreasing}} = 6.05$, 95% CI [5.64, 6.47], $M_{\text{increasing}} = 6.08$, 95% CI [5.63, 6.53]), $t(198) = 0.09$, $p = .926$, $d = 0.01$.

I was also interested in whether efficacy differed depending on whether participants completed the tasks in their preferred order. However there were no main effects and no interactions of condition and order preference on efficacy ratings for either efficacy measure, Fs < 0.44, ps $> .508$, η^2s < 0.07.

Other dependent variables. None of the non-efficacy dependent variables I collected differed by condition. Participants indicated that it would be equally easy to apply for another job, they would apply for their next job equally soon, and they believed their next application had an equal chance of being successful, ts < 0.73, ps $> .469$, ds < 0.10. Participants in the *increasing-difficulty* condition ($M = 5.37$, 95% CI [4.95, 5.80]) did think their application represented them marginally better than those in the *decreasing-difficulty* condition ($M = 4.85$, 95% CI [4.44, 5.26]), but this effect was not significant, $t(198) = 1.71$, $p = .088$, $d = 0.24$.

Performance. My blind coders did not have high reliability in their performance ratings ($\alpha = 0.65$). However neither the ratings of each individual coder, nor the combined average of their ratings, showed a difference in performance by condition, ts < 0.70, ps > .487, ds < 0.10, so I feel confident that there was no difference in performance by condition despite the low reliability of their ratings.

Awareness of changing task difficulty. Those in the *increasing-difficulty* condition indicated that the tasks were getting harder ($M = 64.86$, 95% CI [61.22, 68.50]), whereas those in the *decreasing-difficulty* condition indicated that the tasks were getting easier ($M = 32.56$, 95% CI [28.37, 36.75]), and these assessments of changing difficulty were different from one another, $t(198) = 11.55$, $p < .001$, $d = 1.64$, confirming that my manipulation was successful.

Discussion

Because the paradigm in this experiment diverges so greatly from the paradigms in which I do find an effect of task ordering on efficacy, there are any number of reasons that effect may not occur in this experiment. As hypothesized in Experiment 7, perhaps participants need to consider and report their efficacy more than once in order to recognize the trajectory of their confidence and skill. Applying to jobs also differs greatly from all my previous tasks in that there is no correct answer. Perhaps feedback is a crucial part of the story, or perhaps applying for jobs simply does not create efficacy in the same way that a task with a correct answer does. I will return to this question more in the discussion section for this chapter.

It is worth noting again, however, that these results still do not conform to participants' predictions from Part 1, in which participants predicted that completing tasks in increasing-difficulty order would create more felt efficacy than completing them in decreasing-difficulty order. Despite the fact that participants in Experiment 3 made this prediction about job application tasks specifically, the current experiment found no effect of task ordering on reported efficacy.

Part 3: Correcting Mispredictions

Across eight experiments, I have first shown that people predict that completing tasks in increasing-difficulty order will create more efficacy than completing them in decreasing-difficulty order (Experiments 1–3). I then show that these predictions are inaccurate. Either the opposite is true, and completing tasks in decreasing-difficulty order leads to higher ratings of efficacy than completing them in increasing-difficulty order (Experiments 4–6), or there is no effect of order on efficacy (Experiments 7 and 8). In Part 3, I explore a potential mechanism for these mispredictions and investigate what conditions might help participants predict correctly. Experiment 9 thus consists of three different prediction conditions and an experience condition to isolate why people incorrectly predict the effects of task ordering on efficacy and to potentially help participants predict correctly. This design also allows for a direct comparison between predictions and actual reported efficacy.

Experiment 9: Analogy Task, Mispredictions and Correct Predictions

In this experiment, I used the analogy task from Experiments 2 and 5. I hypothesize that people mispredict because they fail to fully extrapolate how each subsequent experience will incrementally change their efficacy. Therefore I modify the prediction conditions to more and more closely mimic the actual experience until participants are able to correctly project their own trajectories.

In the first prediction condition, I directly replicated Experiment 2. In the second prediction condition, I tried to make the experience more similar to the actual task by showing participants all of the analogy rounds in order before asking for their predictions. In the third prediction condition, I showed participants the rounds in order and asked them to predict their efficacy after seeing each round. This condition resembled the experience as closely as possible without actually having participants complete the tasks and receive feedback. It also tested whether reporting efficacy multiple times could be an integral component in the effect of task ordering on efficacy. Finally, I ran a replication of Experiment 5 to offer a direct comparison between predictions and experience.

Method

I pre-registered this experiment on the Open Science Framework (https://osf.io/hx6qz).

Participants. I recruited 502 participants (227 female, 2 gender non-binary, M_{age} = 37.69 years, 95% CI [36.66, 38.72]) through Amazon Mechanical Turk who completed a survey in exchange for $0.90.

Design and task. This experiment uses the analogy task described in Experiments 2 and 5. Participants were randomly placed into one of five between-subject experimental conditions: *prediction 1, prediction 2, prediction 3, experience 1,* or *experience 2*. Each of the three prediction conditions also contains two within-subjects conditions in counterbalanced order: *increasing-difficulty* and *decreasing-difficulty*.

Procedure. *Prediction 1*. The *prediction 1* condition is a nearly direct replication of Experiment 2. Participants were given instructions about the task and a primer on analogies. They were then shown the three rounds (easy, medium, and hard) in a random order. They offered predictions about how they would feel after completing the three rounds of analogies in each increasing-difficulty and decreasing-difficulty order, counterbalanced. Then they answered a few control and demographic questions and finished the survey.

There were only a few minor differences between this procedure and the procedure of Experiment 2. In the current experiment, participants could view the analogies in each round for 5–30 seconds. In Experiment 2 they saw the analogies in each round for exactly 10 seconds. I did not collect preferences in this experiment, and the instructions were a bit more thorough. For the full instructions, please see Appendix 3.

In this condition, I expected to replicate the misprediction from Experiments 1–3 in which participants believe completing tasks in increasing-difficulty order will lead to more efficacy than completing the tasks in decreasing-difficulty order.

Prediction 2. This condition differed from the *prediction 1* condition in only one important way. Participants in this condition saw the rounds in the order they were asked to consider completing them before they answered questions about their anticipated efficacy. They saw all three of the rounds twice, first in one order, then in the other. After each set of three rounds, they answered efficacy prediction questions. Otherwise, this procedure was identical to the *prediction 1* condition procedure.

I expected participants to predict no difference in efficacy by task-order condition in this prediction condition. I thought seeing the rounds in order would somewhat correct predictions from the *prediction 1* condition, since participants will have a more similar experience to those that actually complete the tasks. However, based on the null results from Experiments 7 and 8 in which participants only report their feelings of efficacy one time, I did not expect this change to be enough to fully correct predictions.

Prediction 3. This condition had an identical procedure to the *prediction 2* condition with only one difference. Participants answered predicted efficacy questions after each round of analogies for a total of three predictions per condition instead of only once after all three rounds. In total, participants estimated their efficacy six times, three times in the *increasing-difficulty* condition and three times in the *decreasing-difficulty* condition, which mimics the procedure of the *experience* conditions.

I designed this condition to try to correct the mispredictions in Part 1 and Part 2. Based on the different findings in Experiments 4–6 and Experiments 7 and 8, I believed that reporting feelings of efficacy three times instead of once might allow participants to really hone in on the trajectory of their feelings and correct the misprediction from the original prediction surveys in Part 1.

Experience 1 and 2. The *experience* conditions used an identical procedure to the one I used in Experiment 5. The only difference is that in this experiment, I explicitly labeled the rounds as easy, medium, and hard, and shared previous average scores with participants. I also did not collect any motivation or memory bias measures in this experiment.

I expected the *experience* conditions to replicate my results from Experiments 4–6 in which participants report higher efficacy after completing tasks in decreasing-difficulty order than in increasing-difficulty order.

Perceived efficacy. All participants saw three efficacy questions after all three rounds, with the language modified slightly for experience versus prediction: "After completing the **easy round [hard round]**, then the **medium round**, then the **hard round [easy round]**, how skilled do you [think you would] feel about answering analogies?" (1 = *not at all skilled*, 10 = *very skilled*), with similarly worded questions for both confidence and trust. Participants in the *experience* conditions and in the *prediction 3* condition also answered efficacy questions after each of the first two rounds: "After completing the first [second] round, how skilled do you [think you would] feel about answering analogies?" (1 = *not at all skilled*, 10 = *very skilled*), with similarly worded questions for both confidence and trust.

Attributions for change. In the *prediction 2 and 3* conditions, in which participants saw the rounds in order, I asked "Did you think the analogies were changing in difficulty in your first [second] trial?" on a slider anchored at 0 (*Yes they were getting way easier*) and 100 (*Yes they were getting way harder*) which started at 50 (*No change in difficulty*). In the *experience* conditions, I asked both "Overall, were you getting better at analogies, getting worse, or staying about the same?" and participants moved a slider anchored at 0 (*Getting much worse*) and 100 (*Getting better*) which started at 50 (*Staying the same*). They also answered "Did you think the analogies were changing in difficulty" on a slider anchored at 0 (*Yes they were getting way easier*) and 100 (*Yes they were getting way harder*) which started at 50 (*No change in difficulty*).

Control variables. For all participants, I collected task familiarity: "How familiar are you with analogy tasks similar to the ones you completed today?" (*I have never played a game like that before*, *I have played a game like that a few times*, *I sometimes play games like that*, *I frequently play games like that*, *I play games like that almost every day*), enjoyment: "Overall, how much do you enjoy engaging in these analogy tasks?" (1 = *not at all*, 10 = *very much*), education, income, employment, age, and gender.

Results

Perceived efficacy. *Prediction 1.* Failing to replicate the findings from Experiments 1–3, participants predicted no difference in feelings of efficacy by condition ($M_{\text{decreasing}} = 6.05$, 95% CI [5.63, 6.46], $M_{\text{increasing}} = 5.65$, 95% CI [5.22, 6.08]), $t(202) = 1.33$, $p = .186$, $d = 0.19$. It is possible that seeing the analogies in each round for slightly longer allowed participants to extrapolate their feelings more accurately and required them to rely less on the labeling of round difficulty to make their predictions.

Prediction 2. As expected, participants predicted no difference in feelings of efficacy by condition ($M_{decreasing}$ = 6.03, 95% CI [5.64, 6.42], $M_{increasing}$ = 5.92, 95% CI [5.56, 6.29]), $t(184) = 0.38$, $p = .701$, $d = 0.06$.

Prediction 3. In this condition, participants correctly predicted that completing these analogy tasks in decreasing-difficulty order (M = 4.82, 95% CI [4.35, 5.28]) would lead to greater feelings of efficacy than completing them in increasing-difficulty order (M = 7.24, 95% CI [6.84, 7.65]), $t(202) = 7.78$, $p < .001$, $d = 1.09$. They made this prediction even though they did not answer the analogy questions themselves and were not given any feedback.

Experience. These results replicate my findings from Experiment 5, with participants in the *decreasing-difficulty* condition (M = 5.82, 95% CI [5.39, 6.26]) reporting higher felt efficacy than those in the *increasing-difficulty* condition (M = 4.54, 95% CI [4.13, 4.95]), $t(203) = 4.23$, $p < .001$, $d = 0.59$. This result persists even though I explicitly tell participants that the rounds are easy, medium, and hard difficulty and offer the benchmark of average words found per round.

Comparison. To compare the three prediction conditions to one another, I created a difference score by subtracting the increasing-difficulty prediction from the decreasing-difficulty prediction. There was a difference by condition, $F(2, 294) = 25.15$, $p < .001$, $\eta^2 = 0.15$, in that participants who reported efficacy three times (M = 2.42, 95% CI [1.88, 2.97]) had a much greater difference between task order predictions than those who reported it only once, with either ordered (M = 0.10, 95% CI [-0.35, 0.55]) or random rounds (M = 0.40, 95% CI [,-0.10, 0.90]), ts > 5.45, ps < .001, ds > 0.77. The two null predictions were no different from one another, $t(193) = 0.86$, $p = .389$, $d = 0.12$.

Although asking participants to predict their efficacy three times did correct the misprediction in the sense that participants correctly predicted that completing analogy tasks in decreasing-difficulty order creates more efficacy than completing analogy tasks in increasing-difficulty order, it created a different misprediction. Participants in the *prediction 3* condition accurately predict how completing tasks in increasing-difficulty order make performers feel, $t(213) = 0.89$, $p = .376$, $d = 0.12$, but they grossly overestimate how much efficacy performers feel after completing tasks in decreasing-difficulty order, $t(192) = 4.75$, $p < .001$, $d = 0.69$. Interestingly, in the two prediction conditions in which participants only reported efficacy once, they correctly predicted reported efficacy for completing tasks in decreasing-difficulty order, ts < 0.75, ps > .456, ds < 0.11, but grossly overestimated how much efficacy completing tasks in increasing-difficulty order would create, ts > 3.71, ps < .001, ds > 0.51, as compared to what performers actually reported.

Awareness of changing task difficulty. In both prediction rounds in which I showed participants the rounds in order, I also asked whether they noticed the rounds were changing in difficulty. This serves as both a manipulation check and an attention check, since I explicitly told participants that the rounds were changing in difficulty. Participants in the *prediction 2* condition recognized that the rounds were getting harder in the *increasing-difficulty* condition (M = 78.90, 95% CI [74.34, 83.47]), and easier in the *decreasing-difficulty* condition (M = 31.14, 95% CI [24.79, 37.49]). Participants in the *prediction 3* condition reported similar awareness ($M_{increasing}$ = 79.47, 95% CI [74.75, 84.19], $M_{decreasing}$ = 27.77, 95% CI [21.58, 33.97]).

In the experience condition, I explicitly told participants that the rounds were changing in difficulty, and indeed those in the *increasing-difficulty* condition recognized that the tasks were getting harder (M = 92.63, 95% CI [90.29, 94.96]), whereas those in the *decreasing-difficulty* condition recognized that the tasks were getting easier (M = 14.91, 95% CI [9.61, 20.22]).

Awareness of changing performance. Despite their knowledge that the tasks were changing in difficulty, participants in the experience conditions also attributed their changing scores internally. They reported getting better at analogies in the *decreasing-difficulty* condition (M = 68.45, 95% CI [64.26, 72.63]) and getting worse at analogies in the *increasing-difficulty* condition (M = 24.97, 95% CI [20.45, 29.49]), $t(203) = 13.74$, $p < .001$, $d = 1.93$, even though their overall scores across the three practice rounds were no different ($M_{\text{decreasing}}$ = 10.59, 95% CI [10.22, 10.95], $M_{\text{increasing}}$ = 10.40, 95% CI [10.01, 10.79]), $t(203) = 0.69$, $p = .490$, $d = 0.10$.

Discussion

This experiment gives a direct comparison between predictions about how task ordering affects efficacy and the reality of how task ordering affects efficacy. None of the prediction conditions accurately predicted both the direction and magnitude of the difference in felt efficacy created by completing tasks in increasing-difficulty order versus decreasing-difficulty order. Participants in the *prediction 3* condition who saw the easy, medium, and hard rounds in the order of their predictions and answered efficacy questions after each round came the closest to being correct, accurately predicting that starting with the hardest task and ending with the easiest creates more felt efficacy than completing tasks in the opposite order. However, those participants drastically overestimated exactly how efficacious they would feel after completing the analogy rounds in decreasing-difficulty order. In the other two prediction conditions, participants did not think task ordering would affect feelings of efficacy, which was also incorrect, but closer to the truth than the predictions from Experiment 2, in which participants believed completing the analogy tasks in increasing-difficulty order would lead to more felt efficacy than completing them in decreasing-difficulty order.

While I cannot fully explain the failure to replicate Experiment 2, I do think that the closer the experience of the prediction comes to reality, the more likely participants are to predict feelings of efficacy correctly. In that sense, perhaps even just being able to see the rounds for slightly longer in the current experiment as compared to Experiment 2 allowed those participants in the *prediction 1* condition to complete a few more of the analogies in their head, more closely replicating the experience of people who complete the task and thus creating a slightly more accurate mental simulation.

The *prediction 2* condition, in which participants saw the rounds in order and predicted efficacy only once provides two pieces of new information. First of all, seeing the rounds randomly versus seeing them in order did not seem to change predictions. More interestingly, it seems that reporting efficacy at multiple intervals is fundamentally different from reporting efficacy only after completing all the tasks, which lends further support to the idea that reporting efficacy more than once could moderate the findings in Part 2.

Discussion (Chapter 3)

Every day, people have to make decisions about how to complete various tasks in their lives. The present research shows that when made to choose, people have a strong preference for starting with their easiest tasks and working their way up to their hardest ones. One reason people may prefer to complete tasks in increasing-difficulty order is because they think that order will create a stronger feeling of efficacy than the opposite order. However, the current research exposes that assumption as untrue. In fact, if people are interested in maximizing efficacy, this research suggests that they not only start with their hardest task, but that they also attend to how bad they fell as they struggle through it. After each subsequent task, if they continue to attend to their feelings of efficacy and notice the steady improvement, they will likely end up feeling much better in comparison to starting with their easiest task.

Theoretical contributions

This work contributes to the existing literature in a number of ways. First I provide clear evidence that people's predictions about creating efficacy via the ordering of tasks is incorrect. In partial support of Tracy's (2011) "Eat that Frog!" suggestion, I find that some tasks create greater efficacy when you start with your hardest and end with your easiest than when you start with your easiest and finish with your most difficult. To my knowledge, this is the first empirical evidence that simply reversing the order in which you complete tasks can causally affect feelings of efficacy.

Second, this research adds to the literature linking efficacy to motivation. Even when task ordering affected ratings of efficacy, it did not have any causal impact on motivation. I hypothesize in my introduction that increased efficacy would not affect extrinsic motivation but found that the participants were more extrinsically than intrinsically motivated in these particular tasks. This may explain why I did not find a causal link between task ordering and motivation despite finding high correlations between efficacy, intrinsic motivation, and state motivation.

Third, I explore a potential new mechanism for correcting mispredictions, an important goal, given the growing body of research on the downstream consequences of incorrectly forecasting your own feelings. Affective misprediction can cause mischoice in important contexts such as whether to take a job or move homes (Hsee & Zhang, 2004), whether to seek revenge (Carlsmith, Wilson, & Gilbert, 2008), and whether to take unnecessary pre-emptive action to avoid regret (Gilbert, Morewedge, Risen, & Wilson, 2004), all of which can lead to a decrease in overall happiness. Further, people tend to overweight the impact of external circumstances on happiness, leading to less fulfillment in the long run (Diener & Oishi, 2005) and even negative economic outcomes (Frey & Stutzer, 2014). In this research, I show that the more closely participants replicated the phenomenological experience of completing the tasks, the more accurate their efficacy predictions became. In our daily lives, we do not often make the effort to fully simulate how completing a series of tasks will make us feel. These findings suggest that people may be able to more accurately predict their future psychological states if they spend a bit longer considering exactly how they would feel after each of a series of activities.

Potential Moderators

This chapter did not fully explore potential moderators to the link between task ordering and perceived efficacy, including the type of task, timing of feedback, the performer's expertise, and whether or not the task is novel. In games of chance, for instance, I do not expect that decreasing-difficulty order will lead to more efficacy, since skill, an important component of efficacy, is not involved in the task (see Chapter 2). Similarly with a novel task, people may need to start with the easiest version in order to learn how to complete that type of task before tackling the hardest one.

Another potential moderator is whether participants are looking ahead to the future or looking back into the past. Two studies by Weinstein and Roediger (2010; 2012) suggest that participants are more optimistic and feel better about their past performance when they complete tasks in order of increasing difficulty. The authors manipulated the order in which participants saw general knowledge questions to be easy to hard, hard to easy, or random. They then asked participants to make post-dictions about their performance. Participants in the easy to hard order believed they had answered more questions correctly than did participants in the other two conditions, though actual performance did not differ (2010). Further, they found that participants are more optimistic about their past performance when they saw questions in the easy to hard order and suggest there may be a memory bias in play (2012). This data suggests that there may be benefits to completing tasks in increasing-difficulty order, especially if you are focused on outcomes other than efficacy.

Limitations and future directions

The work presented in this chapter is interesting and important but not complete. This research leaves countless unanswered questions and opportunities for future directions, some of which I will discuss here. To start, there are a number of limitations to the link between task ordering and efficacy, some of which I encountered in this research and some of which require continued exploration. I found that the effect only occurred when participants reported their efficacy multiple times. When reporting efficacy only once, task order did not affect felt efficacy. Perhaps participants need to consider their feelings of efficacy multiple times in order to realize the trajectory that causes this effect, or perhaps a different mechanism underlies this boundary condition. Further research is required to narrow down the situations in which task ordering can impact efficacy.

Second, I have yet to find this effect in tasks that do not have correct answers. I only tested one such task, and I only asked participants to report their efficacy one time, so the nature of the task may not have caused the null effect. Because most day-to-day tasks do not have a correct answer, it would be beneficial to study more tasks which require efficacy but do not have predetermined outcomes for this effect to be most relevant to general task ordering.

Third, the misprediction effect, while important, is not well-defined. It seems to hinge on how closely the instructions for the prediction match the actual experience. When participants walked through the set of tasks exactly as other participants completed it but without actually answering the questions or receiving feedback, they were able to correctly predict which ordering leads to higher efficacy. However, some prediction conditions resulted in null effects and some resulted in the opposite prediction. In order to better help people forecast their own psychological states, it would be worthwhile to further explore the mechanisms contributing to correct versus incorrect predictions about felt efficacy.

General Discussion

Together, these experiments help us understand different ways to generate perceived momentum and perceived efficacy and how those constructs impact performance and performance expectations. In the first chapter, I demonstrate that performers who report gaining momentum expect to perform better than those who report losing or no momentum, but that performance does not actually differ between the two groups. Experiencing momentum may therefore cause actors to miscalibrate their expectations relative to reality, erroneously believing, and even betting, that their performance will be better than the performance of people who are not gaining momentum.

In the second chapter, I use a different paradigm to create a sense of momentum and again find a link between perceived momentum and performance expectations, this time in observers. More importantly, I show that the link between perceived momentum and improved performance expectations only exists when the change in momentum occurs alongside efficacy. When efficacy is attenuated or not present (i.e., games of chance), people predict either no difference in performance or even a reversal in fortunes.

In the third chapter, I explore how people prefer to order tasks, how people believe ordering tasks impacts efficacy, and how ordering tasks actually impacts efficacy. I find that people believe completing tasks in increasing-difficulty order will create greater efficacy than completing them in decreasing-difficulty order and prefer to complete tasks in increasing-difficulty order. In reality, when participants report their efficacy after each set of tasks, decreasing-difficulty order often increases efficacy as compared to increasing-difficulty order. I do not find any instance in which completing tasks in increasing-difficulty order leads to greater reported efficacy than completing them in decreasing-difficulty order as people predict. I am able to at least partly "correct" participants' mispredictions about the effect of task-ordering on their efficacy by helping them to better simulate the experience of completing the tasks in different orders. Specifically, when people report their predicted efficacy three times instead of once, they make more accurate predictions.

Taken together, these three chapters offer a deep dive into both the momentum literature and the efficacy literature and tie them together in unique and previously unexplored ways. In Chapters 1 and 3, I also demonstrate that strongly held beliefs about momentum and efficacy can lead actors to make inaccurate forecasts about such impactful outcomes as their performance or their psychological states. These findings have implications for anyone from competitive athletes at the highest level to people simply attempting to prioritize their daily to-do list.

References

Adler, P. (1981). *Momentum: A theory of social action.* Beverly Hills, CA: Sage Publications.

Albright, C. S. (1993) A Statistical Analysis of Hitting Streaks in Baseball, *Journal of the American Statistical Association, 88*(424), 1175–1183.

Arkes, J. (2011). Do gamblers correctly price momentum in NBA betting markets? *Journal of Prediction Markets*, *5,* 30–52.

Arkes, J. (2013). Misses in "hot hand" research. *Journal of Sports Economics*, *14*(4), 401–410.

Arkes, J., & Martinez, J. (2011). Finally, evidence for a momentum effect in the NBA. *Journal of Quantitative Analysis in Sports*, *7*(3), 1–14.

Attali, Y. (2013). Perceived hotness affects behavior of basketball players and coaches. *Psychological Science, 24*(7), 1151–1156.

Avery, C., & Chevalier, J. (1999). Identifying investor sentiment from price paths: The case of football betting. *The Journal of Business*, *72*(4), 493–521.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*(8), 1369–1378.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117–148.

Bandura, A., Adams, N. E., & Beyer, J. (1977). Cognitive processes mediating behavioral change. *Journal of Personality and Social Psychology, 35*(3), 125–139.

Barling, J., & Beattie, R. (1983). Self-efficacy beliefs and sales performance. *Journal of Organizational Behavior Management, 5*(1), 41–51.

Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353–363.

Briki, W., & Markman, K. D. (2018). Psychological momentum: The phenomenology of goal pursuit. *Social and Personality Psychology Compass, 12*(9), 1–14.

Briki, W., Den Hartigh, R. J. R., Hauw, D., & Gernigon, C. (2012). A qualitative exploration of the psychological contents and dynamics of momentum in sport. I*nternational Journal of Sport Psychology, 43*(5), 365–384.

Briki, W., Den Hartigh, R. J. R., Markman, K. D., & Gernigon, C. (2014). How do supporters perceive positive and negative psychological momentum changes during a simulated cycling competition? *Psychology of Sport and Exercise, 15*(2), 216–221.

Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand". *Psychonomic Bulletin & Review*, *11*(1), 179–184.

Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of Personality and Social Psychology*, *95*(6), 1316–1324.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. Psychological Review, 97(1), 19–35.

Carver, C. S., & Scheier, M. F. (2012). Cybernetic control processes and the self-regulation of behavior. *The Oxford Handbook of Human Motivation*, 28–42.

Ceci, S. J., & Kain, E. L. (1982). Jumping on the bandwagon with the underdog: The impact of attitude polls on polling behavior. *Public Opinion Quarterly, 46*(2), 228–242.

Chau, A. W., & Phillips, J. G. (1995). Effects of perceived control upon wagering and attributions in computer blackjack. *The Journal of General Psychology*, *122*(3), 253–269.

Christophel, D. M. (1990). The relationships among teacher immediacy behaviors, student

motivation, and learning. *Communication Education*, *39*(4), 323–340.
Clotfelter, C. T., & Cook, P. J. (1991). *Selling hope: State lotteries in America*. Cambridge, MA: Harvard University Press.
Clotfelter, C. T., & Cook, P. J. (1993). The "gambler's fallacy" in lottery play. *Management Science*, *39*(12), 1521–1525.
Cornelius, A., Silva III, J. M., Conroy, D. E., & Petersen, G. (1997). The projected performance model: Relating cognitive and performance antecedents of psychological momentum. *Perceptual and Motor Skills, 84*(2), 475–485.
Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, *30*(3), 195–209.
De Castella, K., & Byrne, D. (2015). My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement. *European Journal of Psychology of Education, 30*(3), 245–267.
Den Hartigh, R. J., Gernigon, C., Van Yperen, N. W., Marin, L., & Van Geert, P. L. (2014). How psychological and behavioral team states change during positive and negative momentum. *PloS ONE*, *9*(5), 1–10.
Dhar, R., Huber, J., & Khan, U. (2007). The shopping momentum effect. *Journal of Marketing Research, 44*(3), 370–378.
Diener, E., & Oishi, S. (2005). The nonobvious social psychology of happiness. *Psychological Inquiry*, *16*(4), 162–167.
Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *The American Statistician, 58*(1), 38–45.
Dunn, M. F. (2000). Risk-taking, gambling, speculation, and a behavioral interpretation of market psychology. In *Allied Academies International Conference. Academy for Economics and Economic Education. Proceedings* (Vol. 3, No. 1, pp. 29–33). Jordan Whitney Enterprises, Inc.
Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. East Sussex, England: Psychology Press.
Feltz, D. L., & Lirgg, C. D. (1998). Perceived team and player efficacy in hockey. *Journal of Applied Psychology, 83*(4), 557–564.
Feltz, D. L., & Weiss, M. R. (1982). Developing self-efficacy through sport. *Journal of Physical Education, Recreation & Dance*, *53*(3), 24–36.
Forthofer, R. (1991). Streak shooter—The sequel. *Chance*, *4*(2), 46–48.
Frazier, J. A., & Snyder, E. E. (1991). The underdog concept in sport. *Sociology of Sport Journal, 8*(4), 380–388.
Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401.
Frey, B. S., & Stutzer, A. (2014). Economic consequences of mispredicting utility. *Journal of Happiness Studies*, *15*(4), 937–956.
Gernigon, C., Briki, W., & Eykens, K. (2010). The dynamics of psychological momentum in sport: The role of ongoing history of performance patterns. *Journal of Sport & Exercise Psychology*, *32*(3). 377–400.
Gilbert, D. T., Morewedge, C. K., Risen, J. L., & Wilson, T. D. (2004). Looking forward to looking backward: The misprediction of regret. *Psychological Science*, *15*(5), 346–350.
Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*(5843),

1351–1354.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York, NY: Free Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295–314.

Guenther, C. L., & Kokotajlo, C. (2017). Psychological momentum and risky decision-making. *Journal of Theoretical Social Psychology, 1*(2), 43–51.

Hancock, P. A., Williams, G., Manning, C. M., & Miyake, S. (1995) Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, *5*(1), 63–86. Harris, C. R. (2012). Feelings of dread and intertemporal choice. *Journal of Behavioral Decision Making*, *25*(1), 13–28.

Hartzmark, S. M., & Solomon, D. H. (2012). Efficiency and the disposition effect in NFL prediction markets. *The Quarterly Journal of Finance*, *2*(3), 1250013.

Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences*, *10*(1), 31–37.

Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology*, *86*(5), 680–695.

Hubbard, T. L. (2015). The varieties of momentum-like experience. *Psychological Bulletin, 141*(6), 1081–1119.

Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review*, *39*(1), 25–43.

Iso-Ahola, S. E., & Blanchard, W. J. (1986). Psychological momentum and competitive sport performance: A field study. *Perceptual and Motor Skills*, *62*(3), 763–768.

Iso-Ahola, S. E., & Dotson, C. O. (2014). Psychological momentum: Why success breeds success. *Review of General Psychology, 18*(1), 19–33.

Iso-Ahola, S. E., & Dotson, C. O. (2016). Psychological momentum—A key to continued success. *Frontiers in Psychology, 7*(7).

Iso-Ahola, S. E., & Mobily, K. (1980). "Psychological momentum": A phenomenon and an empirical (unobtrusive) validation of its influence in a competitive sport tournament. *Psychological Reports, 46(*2), 391–401.

Jackson, D., & Mosurski, K. (1997). Heavy defeats in tennis: Psychological momentum or random effect. *Chance*, *10*(2), 27–34.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, 48*(1), 65–91.

Kahneman, D., & Snell, J. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making*, *5*(3), 187–200.

Kardas, M., & O'Brien, E. (2018). Easier Seen Than Done: Merely Watching Others Perform Can Foster an Illusion of Skill Acquisition. *Psychological Science*, *29*(4), 521–536.

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.

Klein, N., & O'Brien, E. (2018). People use less information than they think to make up their minds. *Proceedings of the National Academy of Sciences*, *115*(52), 13222–13227.

Koo, M., & Fishbach, A. (2014). Dynamics of self-regulation: How (un)accomplished goal actions affect motivation. *Journal of Personality and Social Psychology, 94*(2), 183–195.

Ku, G., Malhotra, D., & Murnighan, J. K. (2005). Towards a competitive arousal model of decision-making: A study of auction fever in live and Internet auctions. *Organizational Behavior and Human Decision Processes, 96*(2), 89–103.

Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the “hot hand”. *Chance*, *2*(4), 22–30.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, *97*(387), 666–684.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, *100*(1), 91–108.

Loewenstein, G., & Sicherman, N. (1991). Do workers prefer increasing wage profiles? *Journal of Labor Economics*, *9*(1), 67–84.

Louro, M. J., Pieters, R., & Zeelenberg, M. (2007). Dynamics of multiple-goal pursuit. *Journal of Personality and Social Psychology*, *93*(2), 174–193.

Mace, F. C., Lalli, J. S., Shea, M. C., & Nevin, J. A. (1992). Behavioral momentum in college basketball. *Journal of Applied Behavior Analysis, 25*(3), 657–663.

Mack, M., & Stephens, D. E. (2000). An empirical test of Taylor and Demick's multidimensional model of momentum in sport. *Journal of Sport Behavior, 23*(4), 349–364.

Maglio, S. J., & Polman, E. (2016). Revising probability estimates: Why increasing likelihood means increasing impact. *Journal of Personality and Social Psychology*, *111*(2), 141–158.

Markman, K. D., & Guenther, C. L. (2007). Psychological momentum: Intuitive physics and naive beliefs. *Personality and Social Psychology Bulletin, 33*(6), 800–812.

Metzger, M. A. (1985). Biases in betting: An application of laboratory findings. *Psychological Reports*, *56*(3), 883–888.

Miller, J. B., & Sanjurjo, A. (2014). A cold shower for the hot hand fallacy. IGIER Working Paper No. 518. Available at SSRN: https://ssrn.com/abstract=2450479.

Miller, S., & Weinberg, R. (1991). Perceptions of psychological momentum and their relationship to performance. *Sport Psychologist, 5*(3), 211–222.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517.

Moritz, S. E., Feltz, D. L., Fahrbach, K. R., & Mack, D. E. (2000). The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research Quarterly for Exercise and Sport, 71*(3), 280–294.

Mowrer, O. H. (1960). *Learning theory and behavior*. New York: Wiley.

Nelson, L. D., & Meyvis, T. (2008). Interrupted consumption: Disrupting adaptation to hedonic experiences. *Journal of Marketing Research*, *45*(6), 654–664.

Nevin, J. A. (1969). Signal detection theory and operant behavior: A review of David M. Green and John A. Swets' signal detection theory and psychophysics. *Journal of the Experimental Analysis of Behavior, 12*(3), 475–480.

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, *135*(2), 262–285.

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science, 26*, 784–793.

Perreault, S., Vallerand, R. J., Montgomery, D., & Provencher, P. (1998). Coming from behind:

On the effect of psychological momentum on sport performance. *Journal of Sports & Exercise Psychology*, *20*(4), 421–436.
Pettit, N. C., Sivanathan, N., Gladstone, E., & Marr, J. C. (2013). Rising stars and sinking ships: Consequences of status momentum. *Psychological Science, 24*(8), 1579–1584.
Polisano, C. (2018). Betting against PASPA: Why the federal restrictions on sports gambling are unconstitutional and how they hurt the states. *Jeffrey S. Moorad Sports Law Journal*, *25,* 453–486.
Preacher, K.J. & Hayes, A.F. (2008) *Behavior Research Methods, 40*(3), 879–891.
Raab, M., Gula, B., & Gigerenzer, G. (2012). The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimental Psychology: Applied, 18*(1), 81–94.
Rabin, M., & Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, *77*(2), 730–778.
Ransom, K., & Weinberg, R. (1985). Effect of situation criticality on performance of elite male and female tennis players. *Journal of Sport Behavior, 8*(3), 144–148.
Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.
Schall, T., & Smith, G. (2000). Do baseball players regress toward the mean? *The American Statistician, 54*(4), 231–235.
Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, *26*(3–4), 207–231.
Schunk, D. H. (1995). Self-efficacy and education and instruction. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 281–303). Boston, MA: Springer.
Shaw, J. M., Dzewaltowski, D. A., & McElroy, M. (1992). Self-efficacy and causal attributions as mediators of perceptions of psychological momentum. *Journal of Sport & Exercise Psychology, 14*(2), 134–147.
Shen, L., & Hsee, C. K. (2017). Numerical nudging: Using an accelerating score to enhance performance. *Psychological Science, 28*(8), 1077–1086.
Shermer, M. (2008). Patternicity: Finding meaningful patterns in meaningless noise. *Scientific American*, *299*(5), 48.
Silva III, J. M., Cornelius, A. E., & Finch, L. M. (1992). Psychological momentum and skill performance: A laboratory study. *Journal of Sport & Exercise Psychology, 14*(2), 119–133.
Silva III, J. M., Hardy, C. J., & Crace, R. K. (1988). Analysis of psychological momentum in intercollegiate tennis. *Journal of Sport and Exercise Psychology, 10*(3), 346–354.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.
Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261.
Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94.
Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgement and Decision Making*, *1*(1), 1–12.
Taylor, J., & Demick, A. (1994). A multidimensional model of momentum in sports. *Journal of*

Applied Sport Psychology, 6(1), 51–70.

Tenney, E. R., Logg, J. M., & Moore, D. A. (2015). (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of Personality and Social Psychology, 108*(3), 377–399.

Terrell, D. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty, 8*(3), 309–317.

Tormala, Z. L., Jia, J. S., & Norton, M. I. (2012). The preference for potential. *Journal of Personality and Social Psychology, 103*(4), 567–583.

Touré-Tillery, M., & Fishbach, A. (2014). How to measure motivation: A guide for the experimental social psychologist. *Social and Personality Psychology Compass*, *8*(7), 328–341.

Tracy, B. (2011). *No excuses!: The power of self-discipline.* Boston, MA: Vanguard.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Vallerand, R. J., Colavecchio, P. G., & Pelletier, L. G. (1988). Psychological momentum and performance inferences: A preliminary test of the antecedents-consequences psychological momentum model. *Journal of Sport & Exercise Psychology, 10*(1), 92–108.

Vancouver, J. B., & Purl, J. D. (2017). A computational model of self-efficacy's various effects on performance: Moving the debate forward. *Journal of Applied Psychology, 102*(4), 599–616.

Vandello, J. A., Goldschmied, N. P., & Richards, D. A. (2007). The appeal of the underdog. *Personality and Social Psychology Bulletin, 33*(12), 1603–1616.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*(3), 169–185.

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences, 16*(1), 1–12.

Weinberg, R. S., Richardson, P. A., & Jackson, A. (1981). Effect of situation criticality on tennis performance of males and females. *International Journal of Sport Psychology*, *12*(4), 253–259.

Weinstein, Y., & Roediger, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition, 38*(3), 366–376.

Weinstein, Y., & Roediger, H. L. (2012). The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition*, *40*(5), 727–735.

Williams, J. (2010). Momentum and sports betting. Available at SSRN: https://ssrn.com/abstract=1553150.

Yaari, G., & David, G. (2012). "Hot hand" on strike: Bowling data indicates correlation to recent past results, not causality. *PloS One, 7*(1), 1–9.

Yaari, G., & Eisenmann, S. (2011). The hot (invisible?) hand: Can time sequence patterns of success/failure in sports be modeled as repeated random independent trials? *PloS One*, *6*(10), 1–10.

Zhu, M., Yang, Y., & Hsee, C. K. (2018). The mere urgency effect. *Journal of Consumer Research*.

Appendix 1

Previous experiments used to determine sample size for Experiments 1a–d

I conducted three prior experiments in which observers viewed the past performance rankings of two players or teams (e.g., tennis, basketball) who would actually compete in the future. Observers assessed the players' momentum moving into the competition and made predictions about the outcomes. I manipulated whether observers believed that players' rankings were staying the same (i.e., static), the favorite was falling in rankings, the underdog was rising in rankings, or both players were simultaneously moving in the rankings (i.e., favorite falling, underdog rising). When prior rankings were static, the underdog was judged to have significantly less momentum than when both players were changing rankings (*d*s = 1.00, 0.41, and 1.13 for the three experiments run). I used these effect sizes to determine the sample sizes for the experiments in this paper.

Pretest to select tasks for Experiments 1a–d

Procedure. To identify a hard, medium, and easy set of letters for the word-find tasks, I recruited 307 participants (194 female, 1 gender non-binary, M_{age} = 35.42 years, 95% CI [34.13, 36.71]) on Amazon Mechanical Turk in exchange for $0.20 and asked them to create words from one of six sets of letters (aiming for approximately 50 participants per set). I created these letter sets on the basis of my own experience playing similar word games. The letter sets were: AEODTSNCPYRK, AEITFMNLPRYG, EIOBTJNCMYRP, AEODTHNCPYRK, AEIDTSNHPYRM, and AEIBTJNCKYDH. Participants for these pretests saw the same instructions and attention check items that participants saw in the main experiments and worked on one of the word-find tasks for 1.5 minutes.

Scores. The average score on each of the six letter sets was, respectively: 14.98, 15.10, 8.08, 13.00, 14.23, and 10.60. I selected the hardest (M = 8.08, 95% CI [6.76, 9.40]), easiest (M = 15.10, 95% CI [13.00, 17.19]), and median-difficulty sets (M = 10.60, 95% CI [9.09, 12.11]) to use as stimuli in Experiments 1a–d. The average completion rates for each of these sets were statistically different from one another, ts > 2.53, ps < .013, ds > 0.51. Finally, for the competition I chose the second easiest set of letters (M = 14.98, 95% CI [12.38, 17.58]) in order to make the task more enjoyable for participants.

Results including all participants

Experiments 1a–d *(N =1,822)*. The effect of experimental condition on perceived momentum ($M_{decreasing}$ = 60.97, 95% CI [59.42, 62.53]; $M_{increasing}$ = 42.31, 95% CI [40.64, 43.98]), $t(1820)$ = 16.04, p < .001, d = 0.75, and expected likelihood to win ($M_{decreasing}$ = 5.32, 95% CI [5.16, 5.48]; $M_{increasing}$ = 4.76, 95% CI [4.61, 4.91]), $t(1820)$ = 4.97, p < .001, d = 0.23, was statistically significant; the effect on performance was non-significant ($M_{decreasing}$ = 20.51, 95% CI [19.65, 21.38]; $M_{increasing}$ = 19.78, 95% CI [18.95, 20.61]), $t(1820)$ = 1.21, p = .227, d = 0.06.

Experiment 2 *(N = 302)*. The effect of experimental condition on perceived momentum ($M_{decreasing}$ = 18.33, 95% CI [14.94, 21.72]; $M_{increasing}$ = -1.96, 95% CI [-6.11, 2.19]), $t(300)$ = 7.50, p < .001, d = 0.87, and expected likelihood to win ($M_{decreasing}$ = 5.63, 95% CI [5.27, 6.00]; $M_{increasing}$ = 4.57, 95% CI [4.18, 4.96]), $t(300)$ = 3.91, p < .001, d = 0.45, was statistically significant; the effect on performance was non-significant ($M_{Decreasing}$ = 11.64, 95% CI [10.40, 12.87]; $M_{increasing}$ = 11.65, 95% CI [10.37, 12.92]), $t(300)$ = 0.01, p = .992, d < 0.01.

Experiment 3 *(N = 409)*. The effect of experimental condition on perceived momentum was statistically significant ($M_{decreasing}$ = 61.22, 95% CI [57.68, 64.77]; $M_{increasing}$ = 44.47, 95% CI [40.50, 48.44]), $t(407)$ = 6.20, p < .001, d = 0.61. A 2 × 2 ANOVA conducted on performance expectations indicated that participants in the *decreasing-difficulty* condition were more likely to believe that they were going to win (M = 6.13, 95% CI [5.80, 6.46]) than were participants in the *increasing-difficulty* condition (M = 5.95, 95% CI [5.63, 6.27]), $F(1, 405) = 10.06$, $p = .002$, $\eta^2 =$ 0.02. Unsurprisingly, *favorites* also believed that they were going to be more likely to win (M = 6.10, 95% CI [5.77, 6.44]) than did *underdogs* (M = 5.17, 95% CI [4.87, 5.48]), $F(1, 405) =$ 19.28, $p < .001$, $\eta^2 = 0.04$, but there was no interaction, $F(1, 405) = 0.85$, $p = .357$, $\eta^2 < 0.01$. The experimental condition did not affect actual performance levels ($M_{\text{inc-fav}}$ = 10.23, 95% CI [8.75, 11.71], $M_{\text{inc-under}}$ = 9.41, 95% CI [7.22, 11.61], $M_{\text{dec-fav}}$ = 11.57, 95% CI [9.70, 13.45], $M_{\text{dec-under}}$ = 9.82, 95% CI [8.07, 11.57]) after controlling for average practice scores, $F(1, 404) = 0.11$, $p =$ .741, $\eta^2 < 0.01$, but participants in the *favorite* condition scored higher than those in the *underdog* condition, $F(1, 404) = 4.28$, $p = .039$, $\eta^2 < 0.01$. There was no interaction, $F(1, 404) = 1.23$, $p =$ .268, $\eta^2 < 0.01$.

Analysis of practice round scores

Experiments 1a–d. Average practice scores were marginally higher in the *decreasing-difficulty* condition (M = 11.88, 95% CI [11.53, 12.22]) from the *increasing-difficulty* condition (M = 11.45, 95% CI [11.13, 11.76]), $t(1384) = 1.80$, $p = .072$, $d = 0.10$.

Experiment 2. Average practice scores were no different in the *decreasing-difficulty* condition (M = 10.91, 95% CI [10.02, 11.80]) from the *increasing-difficulty* condition (M = 11.24, 95% CI [10.20, 12.29]), $t(258) = 0.48$, $p = .632$, $d = 0.06$.

Experiment 3. Average practice scores were higher in the *decreasing-difficulty* condition (M = 12.26, 95% CI [11.52, 13.01]) than in the *increasing-difficulty* condition (M = 11.10, 95% CI [10.40, 11.79]), $t(342) = 2.27$, $p = .024$, $d = 0.25$. This difference was driven solely by a difference in the number of words found on their easy set ($M_{\text{decreasing}}$ = 15.88, 95% CI [14.83, 16.93]; $M_{\text{increasing}}$ = 12.13, 95% CI [11.29, 12.98]), $t(342) = 5.53$, $p < .001$, $d = 0.60$. The averages of the medium and hard sets of letters were the same across conditions, ts < 1.40, ps > 0.162, ds < 0.16. Because the participants in the *decreasing-difficulty* condition saw the easy set of letters last, I believe this difference is a practice effect and not a skill effect.

Analysis of self-efficacy

Experiments 1a–d. Participants' self-efficacy was no different prior to completing the practice rounds in each condition, ($M_{\text{decreasing}}$ = 6.37, 95% CI [6.21, 6.53], $M_{\text{increasing}}$ = 6.48, 95% CI [6.31, 6.65]), $t(1147) = 0.90$, $p = .370$, $d = 0.05$, but started to diverge immediately after the first practice round. Because participants in the *decreasing-difficulty* condition saw the hard set of letters first, they reported lower self-efficacy levels (M = 4.82, 95% CI [4.64, 5.01]) than those in the *increasing-difficulty* condition (M = 5.93, 95% CI [5.75, 6.10]), $t(1147) = 8.44$, $p > .001$, d = 0.50. Just prior to the competition, however, participants' self-efficacy was significantly higher in the *decreasing-difficulty* condition (M = 6.56, 95% CI [6.40, 6.73]) than in the *increasing-difficulty* condition (M = 5.01, 95% CI [4.82, 5.20]), $t(1147) = 12.24$, $p < .001$, $d = 0.72$.

Experiment 2. Participants' self-efficacy was no different prior to completing the practice rounds in each condition, ($M_{\text{decreasing}}$ = 6.39, 95% CI [6.02, 6.77], $M_{\text{increasing}}$ = 6.37, 95% CI [6.01, 6.73]), $t(258) = 0.09$, $p = .928$, $d = 0.01$, but started to diverge immediately after the first practice round. Because participants in the *decreasing-difficulty* condition saw the hard set of letters first, they reported lower self-efficacy levels (M = 4.84, 95% CI [4.45, 5.24]) than those in the *increasing-difficulty* condition (M = 6.05, 95% CI [5.70, 6.39]), $t(258) = 4.53$, $p > .001$, $d = 0.56$. Just prior to the competition, however, participants' self-efficacy was significantly higher in the *decreasing-difficulty* condition (M = 6.56, 95% CI [6.16, 6.96]) than in the *increasing-difficulty* condition (M = 4.91, 95% CI [4.50, 5.33]), $t(258) = 5.61$, $p < .001$, $d = 0.70$.

Experiment 2 analysis with control variables included

Table A1.1 *Summary of Hierarchical Regression Analysis Predicting Performance Expectations, Accuracy, and Competition Performance*

	Model 1: Performance Expectations			Model 2: Accuracy			Model 3: Competition Performance		
Variable	*B*	*SE B*	*t*	*B*	*SE B*	*t*	*B*	*SE B*	*t*
Experimental condition	0.89	0.25	3.52***	0.85	0.40	2.12*	0.16	0.65	0.24
Average practice score	0.12	0.02	5.21***	0.08	0.04	2.16*	0.91	0.06	14.79***
Age	0.01	0.01	0.69	-0.06	0.02	-3.49***	-0.05	0.02	-1.83
Gender	-0.27	0.26	-1.04	-1.39	0.41	-3.43***	-1.32	0.66	-2.00*
Education	0.02	0.19	0.12	0.34	0.30	1.15	0.52	0.49	1.07
Familiarity with task	-0.02	0.15	-0.12	-0.09	0.24	-0.38	0.89	0.39	2.29*
Fun	0.53	0.08	6.30***	0.27	0.13	2.05*	0.51	0.22	2.36*
Adjusted R^2	0.29			0.12			0.53		
N	260			260			260		

Experimental condition is coded as 0 = *increasing-difficulty*; 1 = *decreasing-difficulty*.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$

Experiment 2 deviation from pre-registration

In the process of writing the paper, I changed the labels of my experimental conditions (to increasing-difficulty and decreasing-difficulty), so the labels in the pre-registration no longer align with the labels in the paper. I also arrived at the idea to conduct a signal detection analysis to quantify miscalibration after the study was completed (thanks to reviewer feedback); therefore, this analysis was not pre-registered.

Additional items

Attention check questions. I asked several attention check questions to make sure participants had read and understood the rules: “How many practice rounds of finding words will you do today?” (*1 round*, *2 rounds*, *3 rounds*), “How long will you have to find words during each round?” (*1 minute*, *1.5 minutes*, *2 minutes*), “Can you repeat the same word?” (*Yes*, *No*), “Can you repeat a letter within a word?” (*Yes*, *No*) and “What is the minimum number of letters for each word?” (*2 letters*, *3 letters*, *4 letters*).

Other items in surveys. Most of my surveys conclude with three exploratory text box questions: “Do you think your performance in the practice rounds affected your competition performance? If so, how?”, “Do you have any thoughts about the competition?”, and “Do you have any other thoughts about the experiment?”

Appendix 2

Experiment 1 Supplementary Materials

Manipulation graphics

Static.

Raonic is currently ranked **14th** in the world.

Simon is currently ranked **15th** in the world.

12	-		Kevin Anderson	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	-		Marin Cilic	27	2,405	23	0	0
14	-		Milos Raonic	25	2,270	20	360	0
15	-		Gilles Simon	31	2,145	24	90	45
16	-		David Goffin	25	1,835	25	45	45
17	-		Bernard Tomic	23	1,720	29	180	45

Favorite falling.

In the past year, Raonic has **fallen** from **8th** in the world to **14th** in the world.

Simon is currently ranked **15th** in the world.

12	-		Kevin Anderson	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	-		Marin Cilic	27	2,405	23	0	0
14	-		Milos Raonic	25	2,270	20	360	0
15	-		Gilles Simon	31	2,145	24	90	45
16	-		David Goffin	25	1,835	25	45	45
17	-		Bernard Tomic	23	1,720	29	180	45

Underdog rising.

Raonic is currently ranked **14th** in the world.

In the past year, Simon has **risen** from **21st** in the world to **15th** in the world.

13	-		Marin Cilic	27	2,405	23	0	0
14	-		Milos Raonic	25	2,270	20	360	0
15	-		Gilles Simon	31	2,145	24	90	45
16	-		David Goffin	25	1,835	25	45	45
17	-		Bernard Tomic	23	1,720	29	180	45

Both moving.

In the past year, Raonic has **fallen** from **8th** in the world to **14th** in the world.

In that same time, Simon has **risen** from **21st** in the world to **15th** in the world.

12	-		Kevin Anderson	29	2,475	24	180	45
13	-		Marin Cilic	27	2,405	23	0	0
14	-		Milos Raonic	25	2,270	20	360	0
15	-		Gilles Simon	31	2,145	24	90	45
16	-		David Goffin	25	1,835	25	45	45
17	-		Bernard Tomic	23	1,720	29	180	45

Attention check

For an attention check, participants "enter[ed] each player's correct ranking into the space below...(e.g., 14)" to ensure that they paid attention to the information I provided. Participants reported each player's current ranking, and depending on their conditions, some participants also reported each player's previous rankings.

Exploratory results

Players' expectation to win. Participants' beliefs about who they thought the players expected to win showed some interesting patterns. For these questions, 0 represents the player himself and 1 represents the opponent, so the closer to zero the number is, the more people think the player believes in himself.

Underdog's expectation to win. In the *both static* condition, 88% of participants think the underdog thinks he will win (M = 0.12, 95% CI [0.06, 0.19]), and that percentage increases to 97% (M = 0.03, 95% CI [0.00, 0.06]) in the *both moving* condition. Those are different, t(207) = 2.57, p = .011, d = 0.36, and the *underdog rising* prediction (M = 0.05, 95% CI [0.01, 0.09]) is slightly different from the *both static* condition as well, t(210) = 1.98, p = 0.049, d = 0.27. The *favorite falling* condition was no different from any of the others (M = 0.06, 95% CI [0.02, 0.11]), ts < 1.50, ps > 0.14, ds < 0.20 and *underdog rising* and *both moving* are the same as well, $t(207) = 0.68$, $p = .499$, $d = 0.09$.

Favorite's expectation to win. The favorite showed a more interesting pattern in that participants did not think the favorite believed in himself when he had been falling through the rankings. When participants did not have information about the favorite falling, in the *both static* ($M = 0.05$, 95% CI [0.01, 0.09]) and *underdog rising* ($M = 0.06$, 95% CI [0.01, 0.10]) conditions, 95% and 94% thought the favorite thought he would win, respectively. These are no different from one another, $t(210) = 0.31$, $p = .758$, $d = 0.04$. However, when they knew the favorite had been falling, their projections of his confidence in himself dropped precipitously. In the *favorite falling* condition, only 85% of participants thought the favorite thought he would win ($M = 0.15$, 95% CI [0.08, 0.21]), and in the *both moving* condition it dropped to 81% ($M = 0.19$, 95% CI [0.12, 0.27]). These are not different from one another, $t(211) = 0.95$, $p =.345$, $d = 0.13$, but participants in these two conditions predict that the favorite has much less confidence in himself than in the conditions in which participants do not know he has been falling, ts > 2.17, ps < .031, ds > 0.30.

Want to win. Condition did not affect who the participants wanted to win, $F(421, 3) = 1.78$, $p = 0.15$, $\eta^2 = 0.01$; however, there was a strong preference overall for the underdog, ($M = 0.64$, 95% CI [0.60, 0.69]), *one-sample* $t(424) = 6.23$, $p < .001$, $d = 0.60$.

Experiment 3 Supplementary Materials

Exploratory items

Attention check. Participants completed three questions: "What was Raonic's injury?" (*Wrist, Ankle, None*), "Is Raonic healthy now?" (*Yes, No*), and "What happened to Simon's previous higher-ranked opponents?" (*They were injured, They retired part way through the match, Both of the above*).

Ranking trajectory. After the attention check, I added two more exploratory questions. I reminded participants of the current ranks of the players and of their previous ranks if they were not in the *both static* condition. I then asked, "What do you think Simon will be ranked next year?" and "What do you think Simon will be ranked in two years?" to try to gauge participants' perceptions of each player's trajectory.

Momentum definitions. I also had participants answer the questions "What does "negative (positive) momentum" mean to you? How did you interpret this phrase?"

Exploratory results

Want to win. There was a strong overall preference for the underdog to win in the *no momentum* ($M = 0.63$, 95% CI [0.54, 0.72]) and *momentum with efficacy* conditions ($M = 0.77$, 95% CI [0.68, 0.85]). However, interestingly, in the *momentum without efficacy* condition, participants wanted the favorite to win ($M = 0.39$, 95% CI [0.29, 0.49]), *one-sided* $t(97) = 2.27$, $p = .013$, $d = 0.46$. This may be because they think the favorite deserves a win or because they feel bad for him or because they think the underdog has already benefited enough from his opponents' past misfortunes. This finding lines up with Vandello et al.'s (2007) findings that there must be some injustice for people to root for an underdog and that it often feels as though it is righting a wrong. This is worth investigating further in future research.

Ranking trajectory. I also analyzed my exploratory trajectory measures in which I asked participants to predict the underdog's rank one year and two years in the future (see Figure A2.1). In the *no momentum* condition, participants thought the underdog would improve a little after one year ($M = 13.30$, 95% CI [12.64, 13.96]) and then level off after two years ($M = 14.3$, 95% CI [12.82, 15.83]) and stay around 15th which is the rank at which he started. In the *momentum with efficacy* condition, participants thought his current trajectory would continue and that he would improve after one year to 11th (*M* = 11.19, 95% CI [10.56, 11.82]) and basically hold that rank after two years (*M* = 10.54, 95% CI [8.79, 12.28]). In the *momentum without efficacy* condition, participants do not give the underdog any benefit of the doubt. They think he will maintain his rank of 15 for the next two years (*M*s = 14.49 & 14.68, 95% CIs [13.62, 15.36] & [13.18, 16.19], respectively).

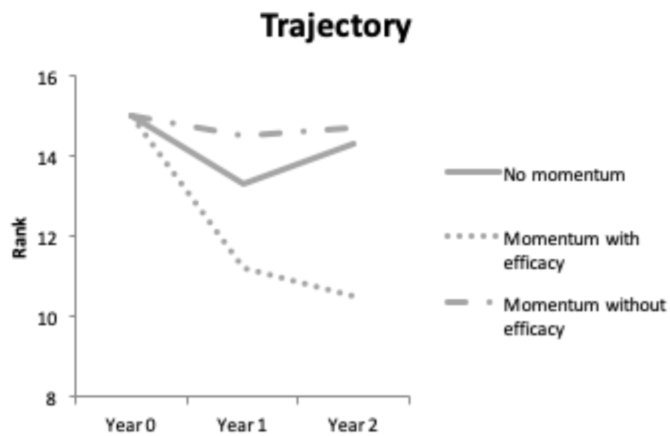


Figure A2.1

Appendix 3

Experiment 1 Supplementary Materials

Exploratory results

Momentum predictions. Participants believed that completing tasks in increasing-difficulty order (M = 21.26, 95% CI [18.28, 24.24]) would create more momentum than completing the tasks in decreasing-difficulty order (M = 11.54, 95% CI [8.00, 15.09]), $t(401)$ = 4.14, $p < .001$, d = 0.41.

Predicted words found. Specific predictions about how many words participants would find in each round did not differ by condition in aggregate ($M_{\text{increasing}}$ = 26.08, 95% CI [24.17, 27.99]; $M_{\text{decreasing}}$ = 25.18, 95% CI [22.72, 27.63]), $t(390)$ = 0.57, p = .566, d = 0.06, or individual round-by-round predictions, ts < .155, ps > .123, ds < 0.16.

Experiment 3 Supplementary Materials

Tasks used

See Experiment 8 supplementary materials for tasks used in this experiment.

Exploratory results

Likely to apply again. Participants indicated that they were more likely to apply to another job after completing the tasks in increasing-difficulty order (M = 7.19, 95% CI [6.80, 7.58]) than in decreasing-difficulty order (M = 6.34, 95% CI [5.88, 6.80]), $t(401)$ = 2.81, p = .005, d = 0.41.

Experiment 5 Supplementary Materials

Attention check questions

I asked several attention check questions to make sure participants had read and understood the rules: "How many rounds will you complete today?" (*Two rounds, Three rounds, Four rounds*); "How many analogies will you complete in each round?" (*6 analogies, 8 analogies, 10 analogies*); "How much is the bonus for each correct question?" (*$0.01, $0.02, $0.05*); and "How many seconds do you have before the page auto-advances and your analogy is considered incorrect?" (*5 seconds, 10 seconds, 15 seconds*).

Analogy pretest

Procedure. I pretested over 150 analogies for this experiment. For the easy round, I selected six analogies that 100 percent of the pretest participants answered correctly so that for a round of six, the expected score was six out of six. For the medium round, I selected six analogies that 50 percent of the pretest participants answered correctly so that the expected score was three out of six. For the hard round, I found six very difficult analogies on the internet and anticipated correct answers would be equal to chance, making the expected score one and a half out of six. I then pretested all three rounds together by recruiting 196 participants (87 female, M_{age} = 34.90 years, 95% CI [33.32, 36.48]) on Amazon Mechanical Turk in exchange for $0.70.

The rounds were all different from one another and similar to my expected value for each round. Participants correctly answered more analogies in the easy round (M = 5.05, 95% CI [4.84, 5.27]) than the medium round (M = 2.75, 95% CI [2.54, 2.96]), $t(390)$ = 14.99, $p < .001$, d = 1.52, and more analogies in the medium round than in the hard round (M = 1.24, 95% CI [1.09, 1.39]), $t(390)$ = 11.44, $p < .001$, d = 1.16, as designed.

Analogies used in Experiment 5

Hard round.

inter : exhume :: piebald : ______

- detour
- homogenous
- heterogenous
- spurn

Sufi : mystic :: eider : ______

- tree
- swan
- swamp
- duck

dense : osmium :: electronegative : ______

- elements
- fluorine
- sodium
- neutron

erstwhile : former :: mithridate : ______

- antidote
- latter
- current
- toxin

effete : fructuous :: chapfallen : ______

- crestfallen
- barren
- effervescent
- precipitous

product : quotient :: exponential : ______

- logarithm
- quadratic
- linear
- constant

Medium round.

flower : butterfly :: dirt : _____

- bugs
- fly
- rats
- sweeper

dog : rabies :: mosquito : ______

- malaria
- death
- plague
- sting

king : throne :: rider: _____

- horse
- chair
- saddle
- seat

ornithologist : birds :: anthropologist : ______

- humankind
- environment
- plants
- animals

major : battalion :: colonel : ______

- command
- army
- regiment
- soldier

hermit : solitude :: intruder : ____

- burglar
- thief
- aim
- privacy

Easy round.

glove : ______ :: monitor : computer

- trounce
- trip
- hand
- trinket

bed : bedroom :: ______ : bathroom

- toilet
- disengage
- pillow
- blanket

car : road :: ship : ______

- prideful
- busy
- shovel
- water

______ : crawling :: birds : flying

- snakes
- cold
- windy
- knapsack

cricket : bat :: hockey : _____

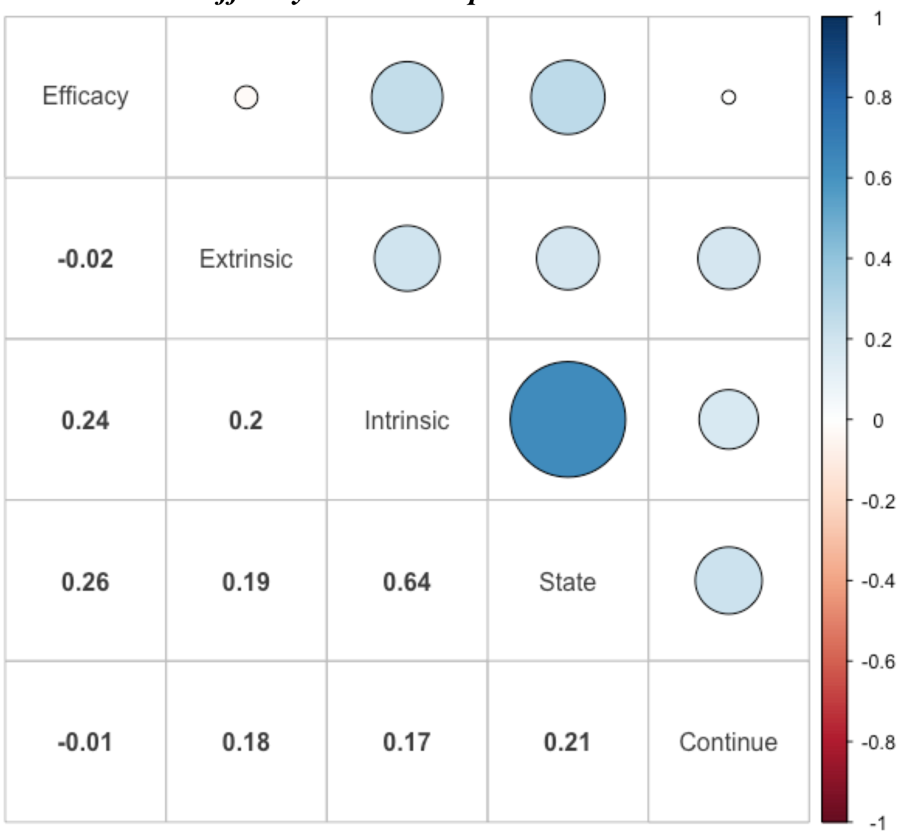
- stick
- player
- ball
- field

train : ______ :: truck : street

- car
- railroad
- busy
- highway

Exploratory results

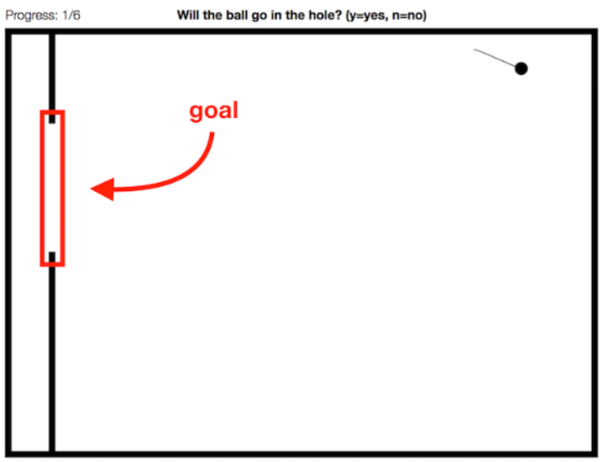
Motivation and efficacy correlation plot.



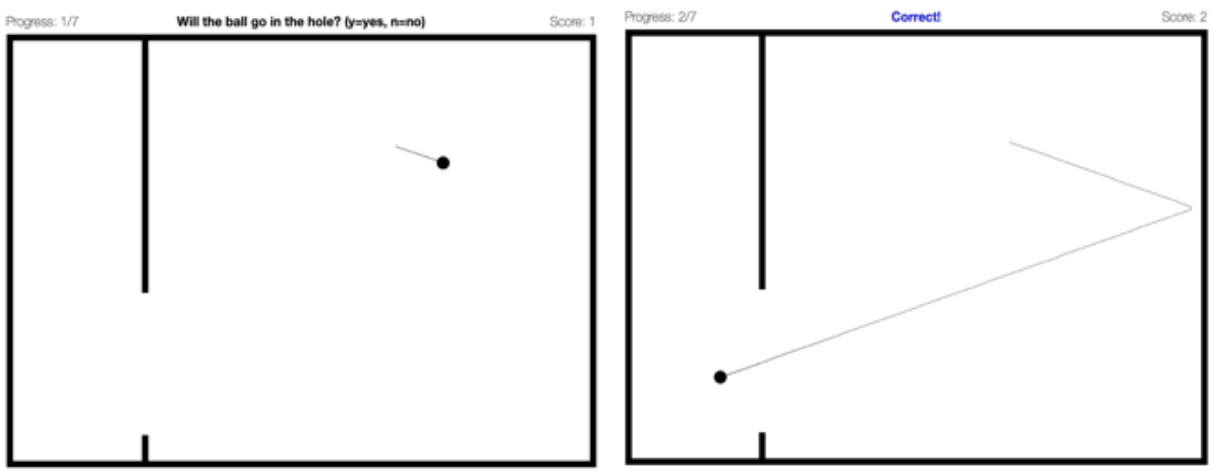
Experiment 6 Supplementary Materials

Task examples

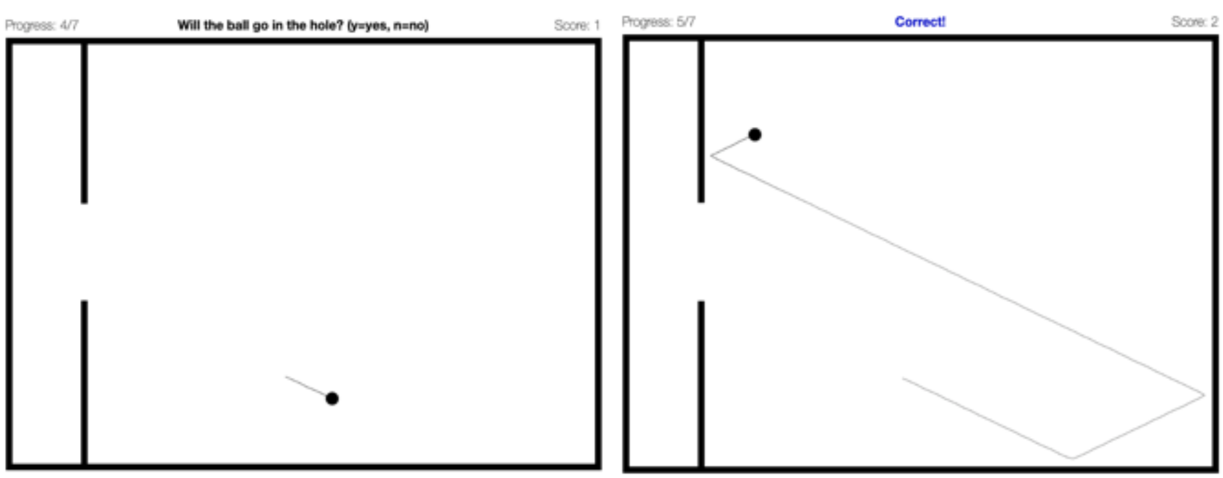
Instructions.



Easy.



Hard.



Stimulus pretest

Procedure. I pretested 52 stimuli for this experiment. I recruited 92 participants (39 female, M_{age} = 35.29 years, 95% CI [33.34, 37.25]) on Amazon Mechanical Turk in exchange for $1.00. Each participant saw 26 of the 52 trials, and each trial was seen by either 46 or 47 different participants. Of those 52 trials, I selected 31 for this experiment: three for the instruction round, seven for each of the three practice rounds, and seven for the competition round. For the easy round, I selected trials with an expected value of 5.75 out of 7 correct. For the medium round, I selected trials with an expected value of 4.34 and for the hard round I selected trials with an expected value of 2.15. I wanted the competition round to be medium difficulty, so I selected trials that had an expected value of 4.17 correct. The trials I used for instruction had an expected value of 2.15 correct out of three. Once I selected the seven trials for each round, I arranged them within their selected round in a random order and then combined the rounds such that I had an ordered list of all 21 trials for the practice rounds. In the *increasing-difficulty* condition I showed participants trials 1–21 broken into three blocks of seven. In the *decreasing-difficulty* condition I showed participants trials 21–1 broken into three blocks of seven. In the *control* condition I showed participants the 21 trials in random order broken into three blocks of seven. The completion rates for the trials in each round are listed below in the order the participants saw the trials.

Increasing difficulty.

Round 1 (Easy). 91.49%, 95.65%, 84.78%, 84.78%, 67.39%, 70.20%, 80.85%

Round 2 (Medium). 57.45%, 68.09%, 71.74%, 59.57%, 50.00%, 60.87%, 65.96%

Round 3 (Hard). 53.19%, 30.43%, 34.78%, 45.65%, 40.42%, 51.06%, 30.43%

Decreasing difficulty.

Round 1 (Hard). 30.43%, 51.06%, 40.42%, 45.65%, 34.78%, 30.43%, 53.19%

Round 2 (Medium). 65.96%, 60.87%, 50.00%, 59.57%, 71.74%, 68.09%, 57.45%

Round 3 (Easy). 80.85%, 70.20%, 67.39%, 84.78%, 84.78%, 95.65%, 91.49%

Instruction trials. 65.96%, 87.23%, 61.70%

Competition trials. 80.85%, 76.09%, 55.32%, 63.04%, 50.00%, 53.19%, 38.30%

Attention check questions

I asked several attention check questions to make sure participants had read and understood the rules: "Which key should you press if you think the ball will bounce off the back wall and then fly through the hole in the center wall?" (*y (yes), n (no)*); "Which key should you press if you think the ball will bounce off the center wall, then the back wall, and then fly through the hole in the center wall?" (*y (yes), n (no)*); "Which key should you hit if you think it is going to bounce twice off the back walls and then go through the hole in the center wall?" (*y (yes), n (no)*); and "How quickly do you need to make your prediction in order to earn extra points?" (*Less than 1 second, Less than 4 seconds, Less than 10 seconds*).

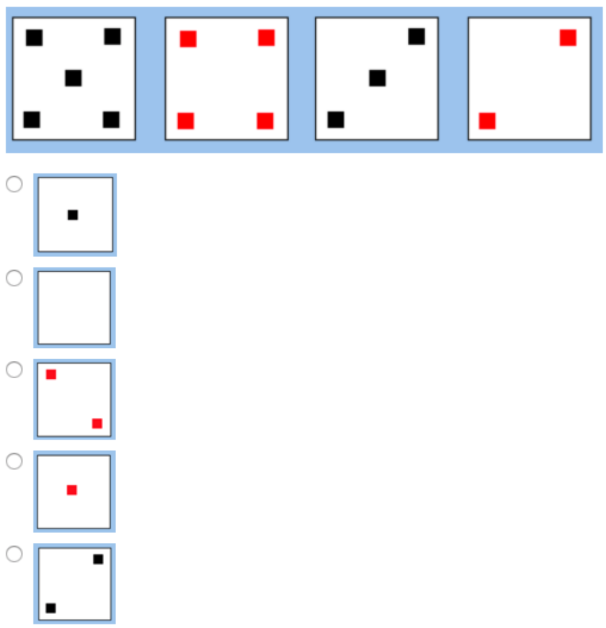
Experiment 7 Supplementary Materials

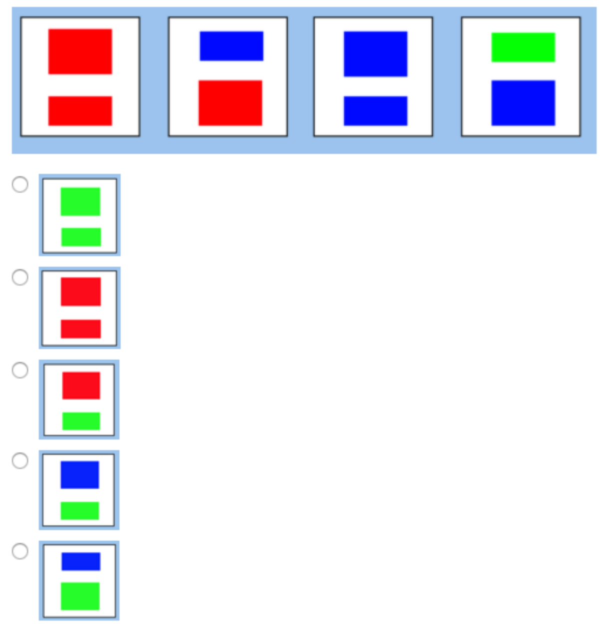
Stimulus pretest

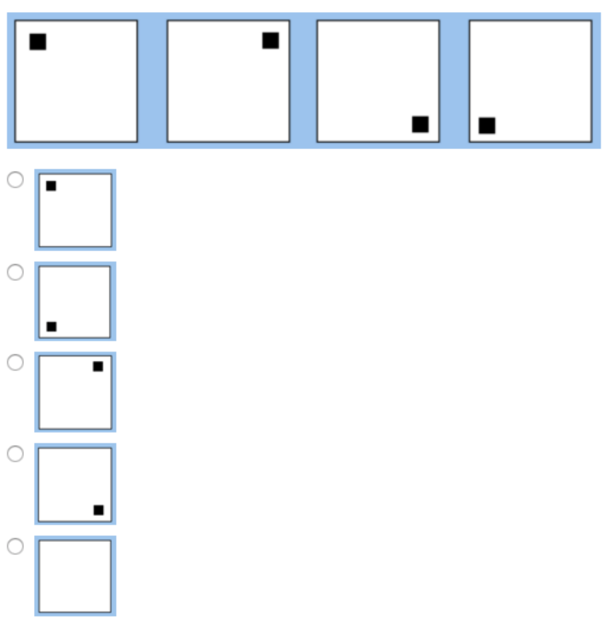
Procedure. I recruited 100 participants (39 female, M_{age} = 35.29 years, 95% CI [33.34, 37.25]) on Amazon Mechanical Turk in exchange for $1.00 with the opportunity to earn a bonus. Each participant saw ten out of twenty of the non-verbal reasoning questions I was testing in a random order. I then chose ten of those questions to use for Experiment 7, choosing a set that progressively changed in difficulty. Based on percentage of participants who answered them correctly, I chose the following ten questions, beginning with the easiest trial and finishing with the most difficult: 86%, 84%, 72%, 69%, 59%, 57%, 48%, 38%, 31%, 24%. There were a few questions that had a lower success rate than 24%, but I wanted to keep the success rate of the most difficult question at chance or above so that participants did not feel as though they were being tricked.

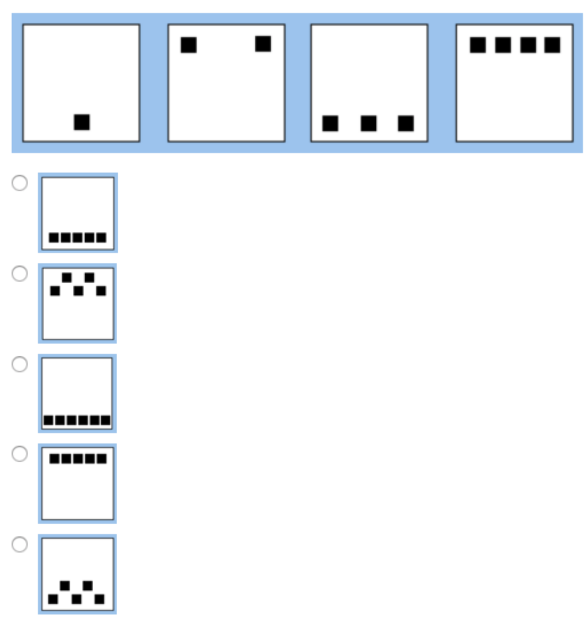
Non-verbal reasoning questions used in Experiment 7

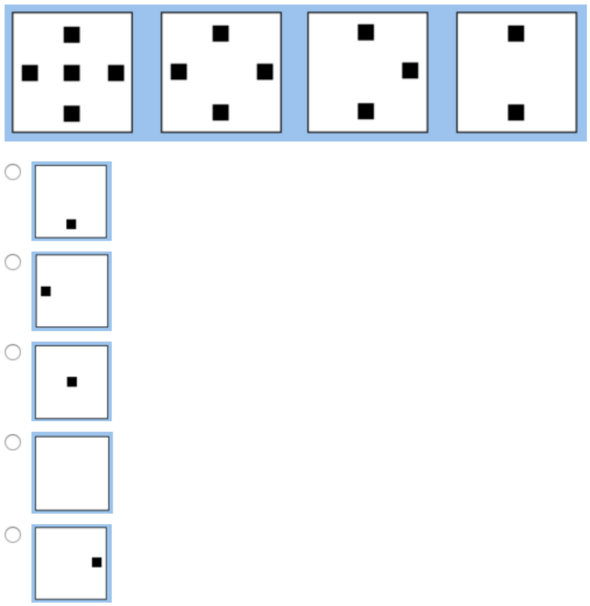
The questions below are displayed in increasing-difficulty order (left to right, top to bottom). In these images, the correct answer is always the top one, but the participants saw the answer choices in a random order.

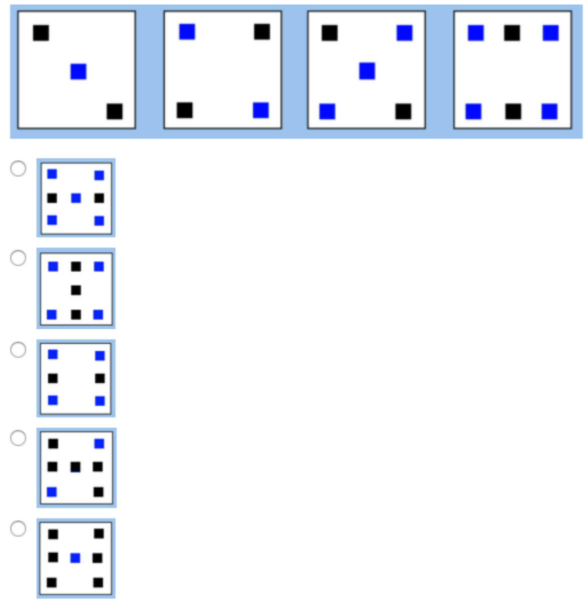


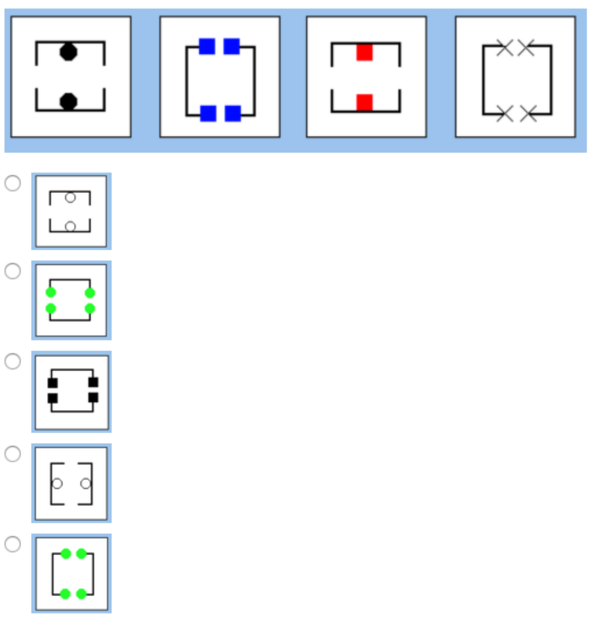


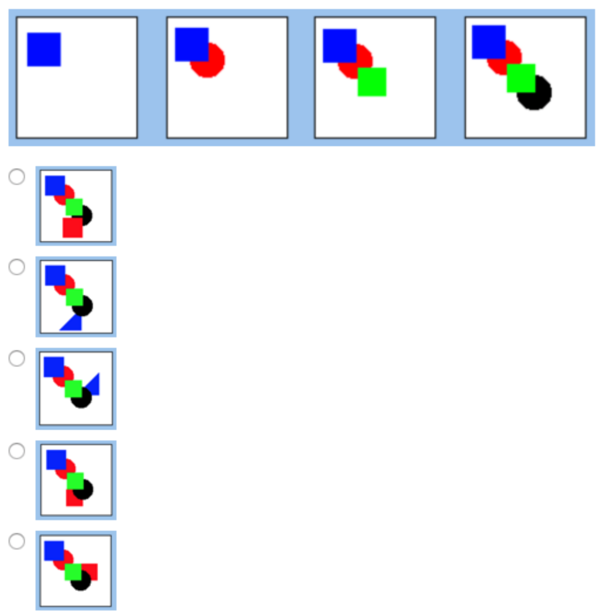


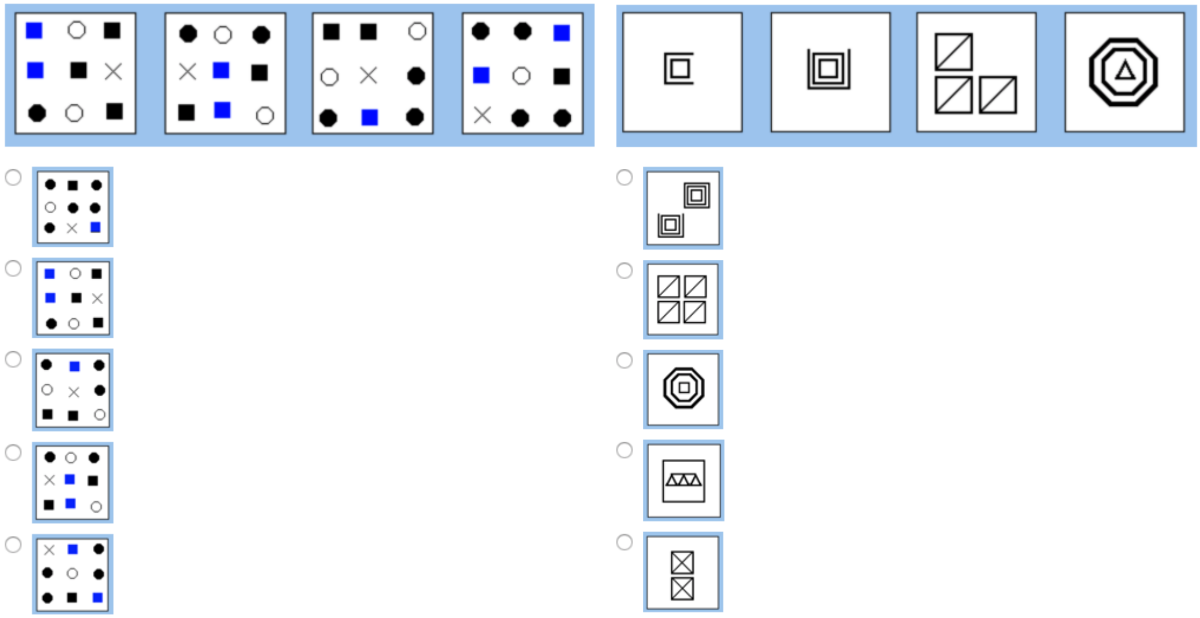






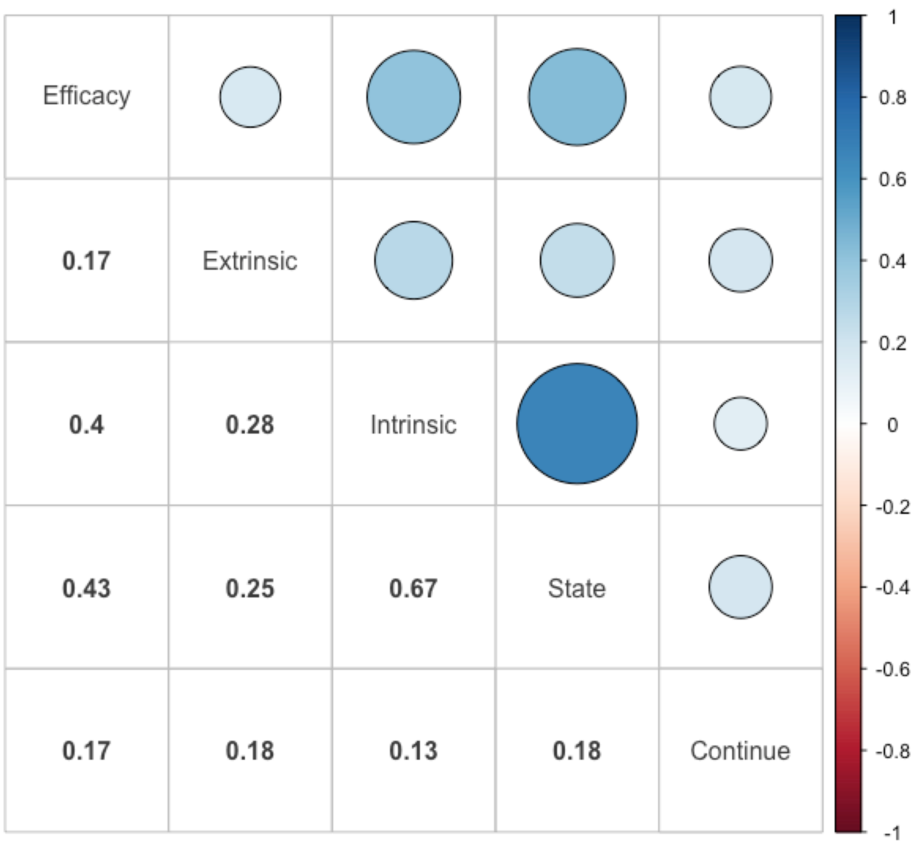






Exploratory results

Motivation and efficacy correlation plot.



Experiment 8 Supplementary Materials

Stimulus pretest

Participants. I recruited 94 first and second year students (68 female, 3 non-binary, M_{age} = 18.74 years, 95% CI [18.60, 18.89]) from the University of California, Berkeley's

Experimental Social Science Laboratory (XLab) through SONA. I chose first and second year students because I was planning to use third and fourth year students for the job application study and wanted to use a similar sample without using the same pool I planned to use for the final study.

Procedure. I showed participants nine different tasks related to applying for jobs and had them rate all nine tasks on the following measures: "How difficult this task is", "How pleasant this task is", "How important this task is in the job application process", "How much you want to complete this task" (1 = *not at all*, 10 = *very*). Participants also used a slider anchored at 0 minutes and 60 minutes with anchors every 10 minutes to indicate how long they thought the task would take to complete. As with the other prediction surveys, participants did not actually complete the tasks.

Tasks used

Task 1. I told participants they would need to fill in a form with the following information about themselves: First name, last name, email address, phone number, emergency contact, emergency contact relationship to applicant, emergency contact phone number.

Task 2. I told participants they would need to fill in a separate form for each of their three most recent jobs with the following information: Employer, position, start date, end date, website.

Task 3. I told participants "In this task, you would provide the contact information of three references, and draft the text of an email you might send to each person asking for him/her to act as a character reference for you as well as notifying him/her that someone may be in touch. You would fill out three forms that look like the one below."

Task 4. I told participants "In this task you would write a short cover letter. It would not need to be more than two paragraphs. You would describe why you want to work in the job you are applying for and what applicable skills you could bring to the position."

Task 5. I told participants "In this task you would be asked to indicate availability for a phone screen with a recruiter. You would need to check your calendar to find times you could take a call and then indicate your availability on a form like the one below." and then showed them a schedule with each weekday and time slots from 10am–11am, 11am–12pm, 1pm–2pm, 2pm–3pm, 3pm–4pm and 4pm–5pm. Each separate day and time slot had a checkbox that someone completing the task could check.

Task 6. I told participants "In this task, you would be asked to describe a challenge you faced or a mistake you made in a previous job and what you learned from the experience."

Task 7. I told participants "In this task you would be asked to create an objective sentence for your resume. An objective sentence usually states what kind of career you are seeking, and what skills and experiences you have that make you ideal for that career." I then offered a couple of examples: "Seeking a position as a clinical practice assistant for a health maintenance organization, utilizing my award-winning writing, research, and leadership skills.; Elementary education teacher looking for a position at a small independent school, where I can apply my five years of teaching experience and my curriculum development skills."

Task 8. I told participants "In this task you would be asked to provide a list of five duties you have executed in prior jobs in bullet point format. You would be asked to make these bullet points as detailed as possible. There are three examples below of what one of your five bullet points could look like." I then offered a few examples: "Served meals, cleared tables, monitored five tables, and provided exceptional customer service to up to 30 customers.; Achieved 100 percent of call performance goals for accuracy, speed, volume, resolution of issues, and customer satisfaction.; Utilized strong interpersonal and communications skills to serve customers; received employee of the month award twice."

Task 9. I told participants "In this task you would be asked to tell us about any clubs and other extracurriculars you participated in during high school. You would list the activity and a sentence or two about your involvement."

Pretest results

I chose five of the nine tasks for the experiment. I wanted to minimize variance in ratings of pleasantness so that all the tasks were similarly aversive. I wanted to maximize variance in perceived difficulty to make the experience as different as possible for those in the *increasing-difficulty* condition as compared to those in the *decreasing-difficulty* condition.

Pleasantness. I selected five tasks that ranged in pleasantness from 4.24 (95% CI [3.82, 4.66]) to 4.80 (95% CI [4.38, 5.22]) which were statistically no different from one another, $t(186) = 1.85$, $p = .066$, $d = 0.27$. Those five tasks are, in increasing-difficulty order: Task 5, Task 2, Task 7, Task 8, Task 6.

Difficulty. Within those five tasks, the perceived difficulty ratings were 2.20 (95% CI [1.89, 2.52]), 2.98 (95% CI [2.59, 3.36]), 4.90 (95% CI [4.53, 5.28]), 4.90 (95% CI [4.49, 5.32]), and 6.02 (95% CI [5.56, 6.48]), respectively. With the obvious exception of the two tasks that were rated the same level of difficulty, these values are all different from one another, ts > 3.10, ps < .002, ds > 0.46.

Importance. Of the five tasks I selected, Task 7 ($M = 6.36$, 95% CI [5.90, 6.83]), the resume objective sentence, was judged to be less important to a job application than the rest of the tasks, ts > 2.10, ps < .038, ds > 0.31, but the other four tasks ranged from means of 7.01 (95% CI [6.61, 7.41]) to 7.63 (95% CI [7.13, 8.13]) and were rated as similarly important to one another, ts < 1.90, ps > .059, ds < 0.28.

Want to complete. Of the five tasks I selected, participants wanted to complete Task 5 ($M = 6.13$, 95% CI [5.62, 6.66]), the job forms, more than the rest of the tasks, ts > 2.58, ps < .011, ds > 0.38, but the other four tasks ranged from means of 4.94 (95% CI [4.50, 5.38]) to 5.22 (95% CI [4.75, 5.70]) and participants indicated no difference in how much they wanted to complete them, ts < 0.88, ps > .380, ds < 0.13.

Tasks used in Experiment 8

I made some minor updates to the tasks between the pretest and the final version of the experiment. The tasks I used are below in their entirety.

Task 1: Scheduling. "Please indicate below any times you would be available next week for a 1 hour phone screen with our recruiter. Please check your calendar to find times you could take a call and then indicate your availability on the form below. For the purposes of this study, it is important that you actually check your calendar and choose accurate time-slots."

	10am-11am	11am-12pm	1pm-2pm	2pm-3pm	3pm-4pm	4pm-5pm	Not available
Monday	☐	☐	☐	☐	☐	☐	☐
Tuesday	☐	☐	☐	☐	☐	☐	☐
Wednesday	☐	☐	☐	☐	☐	☐	☐
Thursday	☐	☐	☐	☐	☐	☐	☐
Friday	☐	☐	☐	☐	☐	☐	☐

Task 2: Job form. "We would like to know about your previous employment. Please provide details about your three most recent jobs. If you have not worked in three jobs, you can provide details about unpaid internships or volunteer positions you have held. Each box needs to be completed, so if you do not remember the exact start and end dates, just make your best guess. If there is no company website or you lack any other requested information, you may enter "n/a" into the box."

Job 1

Employer

Position

Start date

End date

Website

Task 3: Resume objective sentence. "Please create an objective for your resume. An objective usually states what kind of career you are seeking, and what skills and experiences you have that make you ideal for that career. It usually ranges from one to three sentences. Please complete this as though you were actually using it on a job application. In order to make sure you have put adequate thought into this exercise, we require that your answer be at least 100 characters.

Below are a few examples:

Seeking a position as a clinical practice assistant for a health maintenance organization, utilizing my award-winning writing, research, and leadership skills.

Elementary education teacher looking for a position at a small independent school, where I can apply my five years of teaching experience and my curriculum development skills.

Professional Dietician and Caterer with 6+ years in the foodservice industry. Highly entrepreneurial and efficient at building and maintaining client relationships. Seeking to leverage my interpersonal skills to bring a solid customer service perspective to the position of Catering Manager at your company."

Task 4: Job duty descriptions. "Please provide a list of three duties you have executed in prior jobs. Please make these descriptions as detailed as possible. These descriptions are most effective when they include specific metrics. There are examples below of what one of your descriptions of prior duties could look like. In order to make sure that you have put adequate thought into this exercise, we require that each answer be at least 100 characters.

Served meals, cleared tables, monitored five tables, and provided exceptional customer service to up to 30 customers.

Achieved 100 percent of call performance goals for accuracy, speed, volume, resolution of issues, and customer satisfaction.

Utilized strong interpersonal and communications skills to serve customers; received employee of the month award twice."

Task 5: Challenge. "Please describe a challenge you faced or a mistake you made in a previous job and what you learned from the experience. In order to make sure you have answered the question thoroughly, we require that your response be at least 400 characters."

Supplementary results

Ratings of task difficulty. Before completing the tasks, participants rated each task on difficulty, pleasantness, importance, and how much they wanted to complete the task. The table below shows the pre-task ratings for each task on each measure, with 95% confidence intervals and standard errors.

Pre-completion ratings, n=200

		Mean	95% CI Lower	95% CI Upper	Std. Error
Difficult	Interview	2.25	2.01	2.48	0.12
	Previous Jobs	4.36	4.04	4.68	0.16
	Resume Obj	5.45	5.15	5.74	0.15
	Job Duties	4.70	4.41	5.00	0.15
	Challenge	5.67	5.35	6.00	0.16
Pleasant	Interview	4.79	4.47	5.11	0.16
	Previous Jobs	4.04	3.77	4.32	0.14
	Resume Obj	4.26	3.99	4.54	0.14
	Job Duties	4.24	3.97	4.51	0.14
	Challenge	3.82	3.55	4.09	0.14
Important	Interview	7.78	7.47	8.08	0.16
	Previous Jobs	7.50	7.18	7.83	0.17
	Resume Obj	6.70	6.36	7.05	0.18
	Job Duties	7.57	7.29	7.85	0.14
	Challenge	6.89	6.60	7.18	0.15
Want	Interview	6.34	5.97	6.70	0.19
	Previous Jobs	5.11	4.76	5.46	0.18
	Resume Obj	4.79	4.44	5.13	0.17
	Job Duties	4.92	4.59	5.25	0.17
	Challenge	4.45	4.10	4.80	0.18

After completing the tasks, participants rated each of the tasks again on difficulty, pleasantness, and importance. The table below has those post-task ratings for all participants and it also shows how much time participants spent on each task.

Post-completion ratings, n=200		Mean	95% CI Lower	95% CI Upper	Std. Error
Difficult	Interview	1.75	1.57	1.93	0.09
	Previous Jobs	2.70	2.46	2.94	0.12
	Resume Obj	4.55	4.24	4.87	0.16
	Job Duties	5.24	4.93	5.54	0.16
	Challenge	5.41	5.10	5.72	0.16
Pleasant	Interview	5.41	5.10	5.72	0.16
	Previous Jobs	4.59	4.28	4.91	0.16
	Resume Obj	4.55	4.25	4.86	0.15
	Job Duties	3.90	3.62	4.19	0.15
	Challenge	3.84	3.56	4.12	0.14
Important	Interview	7.40	7.07	7.73	0.17
	Previous Jobs	7.03	6.71	7.35	0.16
	Resume Obj	6.57	6.22	6.92	0.18
	Job Duties	7.57	7.29	7.85	0.14
	Challenge	7.04	6.74	7.34	0.15
Time spent	Interview	1.10	1.01	1.20	0.05
	Previous Jobs	2.64	2.47	2.80	0.08
	Resume Obj	2.63	2.35	2.90	0.14
	Job Duties	4.23	3.89	4.58	0.18
	Challenge	4.90	4.49	5.32	0.21

I also broke out the post-task ratings by condition. Below you can see the post-task ratings for the *increasing-difficulty* condition (n = 86).

Post-completion ratings, increasing difficulty condition, n=86		Mean	95% CI Lower	95% CI Upper	Std. Error
Difficult	Interview	1.67	1.40	1.95	0.14
	Previous Jobs	2.85	2.46	3.24	0.20
	Resume Obj	4.56	4.08	5.03	0.24
	Job Duties	5.00	4.56	5.44	0.22
	Challenge	5.85	5.40	6.29	0.22
Pleasant	Interview	4.78	4.28	5.28	0.25
	Previous Jobs	4.36	3.91	4.81	0.23
	Resume Obj	4.65	4.19	5.12	0.23
	Job Duties	3.87	3.46	4.29	0.21
	Challenge	3.87	3.46	4.29	0.21
Important	Interview	7.49	7.00	7.97	0.24
	Previous Jobs	7.05	6.57	7.52	0.24
	Resume Obj	6.56	6.01	7.11	0.28
	Job Duties	7.84	7.43	8.24	0.20
	Challenge	7.23	6.76	7.71	0.24
Time spent	Interview	1.07	0.97	1.16	0.05
	Previous Jobs	2.76	2.48	3.03	0.14
	Resume Obj	2.76	2.28	3.23	0.24
	Job Duties	4.22	3.61	4.83	0.31
	Challenge	4.96	4.28	5.64	0.34

The last table shows the post-task ratings for the *decreasing-difficulty* condition (n = 114).

Post-completion ratings, decreasing difficulty condition, n=114					
		Mean	95% CI Lower	95% CI Upper	Std. Error
Difficult	Interview	1.81	1.56	2.06	0.13
	Previous Jobs	2.59	2.28	2.90	0.16
	Resume Obj	4.55	4.12	4.99	0.22
	Job Duties	5.41	4.98	5.84	0.22
	Challenge	5.08	4.66	5.50	0.21
Pleasant	Interview	5.09	4.62	5.55	0.24
	Previous Jobs	4.77	4.34	5.20	0.22
	Resume Obj	4.48	4.07	4.89	0.21
	Job Duties	3.93	3.53	4.33	0.20
	Challenge	3.93	3.56	4.30	0.19
Important	Interview	7.33	6.89	7.78	0.22
	Previous Jobs	7.02	6.58	7.45	0.22
	Resume Obj	6.58	6.11	7.05	0.24
	Job Duties	7.56	7.18	7.94	0.19
	Challenge	6.89	6.50	7.29	0.20
Time spent	Interview	1.13	0.97	1.29	0.08
	Previous Jobs	2.55	2.35	2.75	0.1
	Resume Obj	2.53	2.2	2.85	0.16
	Job Duties	4.24	3.83	4.65	0.21
	Challenge	4.86	4.33	5.4	0.27

Experiment 9 Supplementary Materials
Prediction 1 instructions

"In today's study, your job is to predict what your experience will be like if you complete the analogies in different orders.

Specifically, you will **imagine** completing the analogies in three different rounds (hard, medium, and easy). **Each round contains 6 analogy questions.**

Here's how today's study will go:

1. You will see all three rounds in a random order.

2. You will imagine completing the analogies in a specific order—either hard, then medium, then easy, or easy, then medium, then hard and predict how that might make you feel. **You will not actually complete the analogies.**

3. You will go through the exercise again, but you will imagine completing the rounds in the opposite order. You will predict how this opposite order might feel different to you.

4. You will answer some questions about yourself and complete the survey.

Please continue to the next page to answer some questions about today's task."