Journal of Experimental Social Psychology 91 (2020) 104032

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

To build efficacy, *eat the frog* first: People misunderstand how the difficulty-ordering of tasks influences efficacy☆

Rachel Habbert, Juliana Schroeder*

University of California, Berkeley, United States of America

ARTICLE INFO

Keywords:
Efficacy
Confidence
Procrastination
Task ordering
Momentum

ABSTRACT

Achieving competency and autonomy in one's life—in other words, being efficacious—is a fundamental human need. A commonly endorsed strategy for building efficacy is summarized by a popular quote: "If it's your job to eat a frog, it's best to do it first thing." The current paper tests this "eat-the-frog-first" strategy, examining whether completing tasks in increasing-easiness order builds efficacy more than increasing-difficulty (or randomized) order. We propose that the eat-the-frog-first strategy does indeed enhance efficacy, but also that people will prefer the opposing order (preferring to complete more difficult tasks later) because they inaccurately believe that doing so will enhance their efficacy. Six experiments and one supplemental experiment (N = 2013) support these hypotheses. In Experiments 1a, 2a, and 3a (predicted efficacy experiments), people believed that completing tasks in increasing-difficulty (vs. increasing-easiness) order would enhance their efficacy, and hence preferred to complete tasks in increasing-difficulty order. But in corresponding Experiments 1b, 2b, and 3b (actual efficacy experiments), completing tasks in increasing-difficulty (vs. increasing-easiness or random) order reduced self-efficacy (or did not meaningfully change it; 3b). We provide evidence in a final study (Experiment 4) that this misunderstanding is due to people simulating the beginning of a sequence (e.g., the struggle of completing the most difficult task) more than the end (e.g., the ease of completing the simplest task). We conclude that people's tendency to delay the difficult incurs unexpected costs to self-worth. To build efficacy, people should start with their hardest task, even though doing so may violate intuition.

> "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."
>
> —*Anonymous* (*a popular aphorism often attributed to Mark Twain*)

Achieving competency and autonomy in one's life—in other words, being efficacious—is a fundamental human need (Deci and Ryan, 1985; Ryan and Deci, 2000). When time and resources allow, humans strive to trust and have confidence in themselves and their skills. One way in which people build self-efficacy is through the successful completion of everyday tasks. Whether giving presentations or competing in sports, for instance, the manner by which skill-based tasks are completed influences how efficacious the performer feels in the task domain. The current paper examines how one element of task completion—the difficulty ordering of tasks—affects people's self-efficacy.

Many popular books theorize about how to build efficacy via task-ordering. For example, the best-selling motivation book "Eat That Frog!" (Tracy, 2007) augments this paper's opening quote to encourage people to "eat the frog" (i.e., do the hardest thing) first during their day, suggesting it will help people to get more done in less time and feel more confident. If this advice is sound, people should complete their hardest task first to enhance efficacy; but instead, people often prefer to delay completing more difficult tasks (e.g., procrastination; Steel, 2007). This seeming discrepancy between folk wisdom (eat the frog first) and descriptive preferences (eat the frog last) suggests that there could be a misalignment in how people *think* that completing tasks will help them to build efficacy and how it actually builds efficacy.

1. Expectations about enhancing efficacy

At least two literatures lead us to hypothesize that people *expect* that starting with their easiest task and working up to the hardest task (i.e., completing tasks in increasing-difficulty order) will help them to enhance efficacy. When trying to augment efficacy, increasing-difficulty order should thus be the preferred ordering of tasks. First, starting with easier tasks and moving toward harder tasks is, of course, the natural

☆ This paper has been recommended for acceptance by Marlone Henderson.

* Corresponding author at: 2220 Piedmont Ave, Berkeley, CA 94609, United States of America.
E-mail address: jschroeder@haas.berkeley.edu (J. Schroeder).

https://doi.org/10.1016/j.jesp.2020.104032
Received 18 March 2020; Received in revised form 21 July 2020; Accepted 24 July 2020

way that people learn new concepts (Mowrer, 1960). In learning math, for instance, one must understand how to count before one can add, and one must understand how to add before one can multiply. Because so much of people's early lives are spent learning via this ordering, it may be a familiar order for them to pursue. Moreover, even for tasks that do not need to be learned (e.g., cleaning one's house), people may overgeneralize learning principles and infer that increasing-difficulty order will enhance self-efficacy. Some studies suggest that people have lay beliefs that efficacy-related concepts such as confidence must "build" via repeated successes (e.g., Feltz and Weiss, 1982; Shaw et al., 1992).

Second, there is a positive relationship between how aversive a task is and how much its completion is delayed (Steel, 2007). In other words, people often procrastinate by delaying aversive tasks. Indeed, a factor analysis of the reasons why people procrastinate found that task aversion, along with fear of failure, were the two main reasons participants delayed starting a task (Solomon and Rothblum, 1984). O'Donoghue and Rabin (2001) further propose that adding more attractive or more important options into an option set can increase procrastination because people forgo less attractive options in hopes of one day completing the more attractive options (which may never get completed). While difficult tasks may not always be aversive (and easy tasks not always attractive), these constructs often go hand-in-hand, suggesting delay of harder tasks.

Although prior findings on procrastination guide us in hypothesizing that people will prefer to complete tasks in increasing-difficulty order, we move beyond the procrastination literature in several ways. For one, the current paper examines not just people's preferences to enhance efficacy but what is actually the best choice for their efficacy. Indeed, anecdotes such as "what doesn't kill you makes you stronger" and motivational self-help books urging people to "eat the frog first" highlight the disconnection between what people may want to do (delay the difficult) and what actually helps them build efficacy (not delay the difficult). This potential disconnects between predictions and reality is the focus of the current paper. Furthermore, the current paper exclusively examines predictions and consequences of task-ordering for self-efficacy, a construct that has not previously been discussed as related to procrastination. Third, as previously mentioned, difficult tasks are not necessarily aversive (e.g., it is possible to have enjoyable difficult tasks); whereas procrastination is delaying the aversive, "frog-eating" (the current paper's focus) is doing the difficult.

A few exceptions to the aforementioned findings have been discovered. For example, sometimes people prefer to start with the hardest or most aversive task to avoid having to dread it (Harris, 2012), or choose to wait for pleasurable events to enhance anticipation (Loewenstein, 1987). Additionally, people sometimes choose to complete even effortful tasks earlier than necessary to free up cognitive resources (Rosenbaum et al., 2014), thus preferring to complete more cognitively demanding tasks before less demanding ones (VonderHaar et al., 2019) ("precrastination"). Such results suggest a preference for increasing-easiness instead of increasing-difficulty order. However, avoiding dread, enhancing anticipation, and freeing cognitive resources have little relevance to building efficacy on a particular task. We thus hypothesize that in the specific instances when people are trying to build efficacy—and not pursuing other goals like minimizing dread—they will prefer increasing-difficulty order because they believe it will enhance their efficacy.

2. Actual enhancement of efficacy

Actual efficacy may be formed differently than expected. Strings of continued successes—which are more likely when a task is getting easier—reliably elicit perceptions of gaining psychological momentum (Iso-Ahola and Dotson, 2014; Markman and Guenther, 2007; Shaw et al., 1992). Psychological momentum has been defined as "the phenomenological experience of goal pursuit" whereby a series of precipitating events evokes a mental simulation of progress (Briki and Markman, 2018, p. 2). For instance, in one study, 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") (Vallerand, Colavecchio, & Pelletier, 1988). This manipulation of momentum contains the same number of wins and losses in both conditions, thereby manipulating only the ordering of the outcomes. In another study, participants who competed against a confederate trained to perform worse (vs. better) in a basketball free throw task, and thus experienced more successes (vs. failures), reported more positive momentum (Shaw et al., 1992).

These findings suggest that completing increasingly easier tasks—thereby experiencing a series of "wins"—will create a sense of psychological momentum in the performer. In turn, momentum is a key input into the experience of efficacy. In a putting study, experienced golfers with positive momentum chose to take a critical putt from further away than those with negative momentum, indicating increased confidence in their putting abilities (Den Hartigh et al., 2017). Similarly, rowers reported higher self-efficacy after gaining positive momentum by winning two races than after developing negative momentum by losing two races (Den Hartigh et al., 2016). In all experiments that measure both momentum and self-efficacy, the two constructs are highly positively correlated (e.g., Attali, 2013; Cornelius et al., 1997; Gilovich et al., 1985; Habbert et al., 2020). Given that completing tasks in increasing-easiness order has been shown to enhance psychological momentum in performers, we therefore predict that completing tasks in increasing-easiness order—compared to increasing-difficulty order or random order—will also build efficacy in performers.

3. Simulating experienced efficacy

Understanding how the trajectory of task-difficulty will influence efficacy may not be psychologically intuitive. In order to predict a future affective state or experience, people must employ mental simulations, but simulations are often "mere cardboard cutouts of reality" (Gilbert and Wilson, 2007, p. 1354). These static representations cannot fully capture the phenomenological nature of the real-time experiences that people attempt to predict. Indeed, people often misunderstand how their own abilities and preferences build and grow over time. For instance, people overestimate their ability to perform complicated skills after repeatedly viewing other people performing them (Kardas and O'Brien, 2018) and believe they will dislike a stimulus (e.g., ice cream, music) after repeated exposure more than they actually do (Kahneman and Snell, 1992). These mispredictions stem from people's failure to accurately simulate multiple experiences in a row, and how much they will learn from or enjoy those experiences, respectively.

One reason for such mispredictions is that people's simulations tend to over-represent the early moments of a future event. As an example of this tendency, people underestimate their happiness levels when imagining being in ill-health—because their simulations focus too much on the initial, and typically worst, moments of these events (Gilbert and Wilson, 2007). When simulating their self-efficacy levels over time, we propose that people will likewise imagine how efficacious they will feel after having completed the first task more than after the final task. In simulating an increasing-difficulty sequence, for example, people will think more about the high efficacy they will feel after completing the first (easiest) task than about the low efficacy they will feel after completing the final (hardest) task. This possibility suggests an intervention to make people more accurate in assessing how task difficulty-order influences their efficacy: having people explicitly predict the impact of the final task they complete (which they may spontaneously tend to ignore).

4. Overview of hypotheses and studies

The current studies ($N = 2013$) test three separate hypotheses. First, people *expect* that completing tasks in increasing-difficulty (vs. increasing-easiness) order will enhance their efficacy and thus prefer to complete tasks in increasing-difficulty order to build efficacy. Second, in contrast to people's expectations, completing tasks in increasing-difficulty (vs. increasing-easiness or random) order will actually *reduce* efficacy. And third, making people focus on the end of their future task trajectory (which they may tend to overlook) helps them recognize the benefits of completing tasks in increasing-easiness order.

The studies test these hypotheses using different task types—verbal skills tasks[1] (completing analogies, Experiment 1, or word finds, Experiment 2) and a more externally valid everyday task (completing job applications, Experiment 3). Experiments 1a, 2a, and 3a examine expectations and preferences; Experiments 1b, 2b, and 3b then test reality. Experiment S1 additionally compares expectations and reality in the same study, testing for a statistical interaction. Because task difficulty can be subjective (e.g., depending on expertise, task familiarity, and other factors; Moore and Healy, 2008), Experiments 3a and 3b allow participants to determine the task difficulty themselves; other experiments operationalize difficulty as tasks that require more effort or have a lower likelihood of success. A final study (Experiment 4) manipulates how much performers mentally simulate their task trajectory, testing not only our hypothesis that those who more fully simulate the end of the trajectory will have more accurate efficacy expectations, but also what people spontaneously simulate when uninstructed. Experiment 4 therefore examines the psychological process by which people estimate their efficacy levels as a function of task difficulty order.

All experiments in this paper are preregistered except Experiment 3a. We report all measures (some in the Supplementary materials), manipulations, and exclusions. Data and materials are publicly available at https://osf.io/puxfq.

5. Experiments 1a & 1b: Building self-efficacy in analogies

Individuals predicted (Experiment 1a) or reported (Experiment 1b) how much efficacy they would or did feel after completing a set of verbal-skills (analogy) tasks in increasing-difficulty or increasing-easiness order. We hypothesized they would expect increasing-difficulty (vs. increasing-easiness) order would enhance efficacy, but that actually increasing-difficulty order would reduce efficacy (compared to increasing-easiness and random order).

5.1. Exp. 1a method

Experiment 1a preregistration can be found on the Open Science Framework (https://osf.io/se6tk).

5.1.1. Participants

Because we did not know what effect size to expect and wanted to be able to identify even a small effect, we predetermined a sample size of 200 within-subjects participants. We 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. A sensitivity power analysis (GPower; Faul et al., 2007)[2] revealed that we have 80% power to detect an effect size of $d = 0.20$ (at $\alpha = 0.05$) using a two-tailed paired t-test.

[1] We decided not to use math tasks because people's math skills vary greatly, which would add unwanted noise into our manipulations.

[2] We use the same sensitivity power analysis tool (GPower) in all of our experiments.

5.1.2. Procedure

We selected a common verbal skills task that has a right or wrong answer and with which most people would have some familiarity—completing analogies. We compiled the analogies from various websites and pre-tested hundreds to determine their difficulty level before choosing the current set (see Supplementary materials for analogies used).

Participants (i.e., predictors) imagined they were participating in a different MTurk survey in which they would "answer three practice rounds of six analogies each." We described to predictors how to complete an analogy:

> Analogy questions take the form "_ 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". A bone is an item dogs like, just as catnip is an item cats like – the two word pairs have the same type of relationship to one another.

We then told them the task and gave them an example:

> You will be given analogies with three items filled in, and you must select the word which completes the analogy from one of four options. Imagine that you answer three separate rounds of six analogies each. For example, if the analogy provided was *orange:fruit::____ :vegetable* and your options were: a) apple, b) rabbit, c) carrot, d) house, you would select c) carrot, completing the analogy *orange:fruit::carrot:vegetable*. An orange has the same relationship to fruit as a carrot has to vegetable—an orange is an item in the broad category fruit, just as a carrot is an item in the broad category vegetable.

We told predictors that "each of the three rounds of analogies 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." To ensure their understanding, we then showed predictors the six analogies in each round for 10 s, displaying the rounds in random order. The information that we gave predictors about the three analogy rounds was based on a real pre-test that we conducted (see Supplementary materials for details), and they viewed the same analogies that participants completed in Experiment 1b.

After learning about the tasks, predictors completed the survey, reporting their preference for task ordering and predicting their efficacy levels after imagining completing the three analogy tasks in increasing-difficulty order and increasing-easiness order (presented in counterbalanced order).

5.1.3. Materials (survey)

5.1.3.1. Efficacy predictions. To collect efficacy predictions, predictors answered three questions, derived from Bandura's (1977, 1988) definition of self-efficacy as "an individual's conviction (or confidence) about his or her abilities to mobilize the motivation, cognitive resources, and courses of action needed to successfully execute a specific task within a given context" (Stajkovic and Luthans, 1998, p. 66). We wanted to keep the scale as simple and face-valid as possible, to ensure that it would measure people's lay intuitions: "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 about answering analogies correctly, after completing all three rounds?", "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 answering analogies correctly, after completing all three rounds?", "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 answer analogies correctly, after completing all three rounds?" (1 = *not at all* [*skilled/confident*], 10 = *very skilled* [*confident/much*]; $\alpha s > 0.92$). We collapsed these

Table 1
Selected examples of participants' reported reasons for choosing increasing-difficulty and increasing-easiness task order in Experiments 1a, 2a, 3a, and 3b.

Experiment number	Participant's preferred task order	"Why did you choose [choice]?"
1a (Analogies)	Increasing-Difficulty Order	Starting with the easier ones will get me warmed up and increase my confidence.
1a (Analogies)	Increasing-Difficulty Order	I always prefer to start out with an easier task, then work my way up to the harder ones
1a (Analogies)	Increasing-Easiness Order	So I can get the hard task out of the way first then enjoy doing the easier tasks.
2a (Word Finds)	Increasing-Difficulty Order	I think doing the easy one will help give me the confidence to be successful in the harder set in the future.
2a (Word Finds)	Increasing-Difficulty Order	Starting off easy will help to build up my confidence.
2a (Word Finds)	Increasing-Easiness Order	Because I want to get the hard one out the way and by the end be able to have the easy one to glide through
3a (Job Applications)	Increasing-Difficulty Order	I like to start off doing easy tasks first, in order to get myself into the mindset of doing work. Once I do so, then my mind becomes adjusted to working and I go into a "work zone" that allows for me to stay focused on any task, regardless of the difficulty.
3a (Job Applications)	Increasing-Difficulty Order	It's more manageable to get started. It's also harder to lose steam and you feel like you're making more progress.
3a (Job Applications)	Increasing-Easiness Order	I feel more motivated getting the difficult tasks out of the way first.
3b (Job Applications)	Increasing-Difficulty Order	Because if it starts out easy then I have motivation to keep going and actually finish the application. If it is hard initially I might feel intimidated and decide not to continue the application.
3b (Job Applications)	Increasing-Difficulty Order	I psychologically feel like I will have accomplished more in a shorter period of time.
3b (Job Applications)	Increasing-Easiness Order	Because I want to get the most difficult and thought-provoking stuff out of the way before I get mentally tired. That way I can just ride that momentum forward and breeze through the easy tasks.

Note. We selected three examples from each study in which we asked participants to defend their choice. Because increasing-difficulty order was more commonly picked than increasing-easiness order in each study, we included two examples for increasing-difficulty order and one example for increasing-easiness order from each study.

three questions into one efficacy index for analysis.

5.1.3.2. Order preferences. To measure predictors' task order preferences (specifically, their intuition about the best way to build efficacy), we told them, "Your goal is to feel the most confident and the most skillful after completing all three rounds of these analogies" and then asked "To achieve this goal, in which order would you prefer to see the practice rounds?" (Easy, then medium, then hard; Hard, then medium, then easy). For exploratory analysis, we also asked "Why did you choose [choice]?" Examples of participants' responses can be found in Table 1 and are also posted in full on OSF.

5.1.3.3. Control measures. We collected predictors' age, education, and gender, and measured their experience with the task: "How familiar are you with analogy tasks similar to the one described 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).

5.2. Exp. 1a results

As hypothesized, predictors believed that completing tasks in increasing-difficulty order ($M = 6.27$, 95% CI [6.01, 6.57]) would create more efficacy than completing tasks in increasing-easiness order ($M = 5.75$, 95% CI [5.43, 6.06]), *paired* $t(199) = 2.60$, $p = .010$, $d = 0.19$. Consistently, more predictors preferred to complete the tasks in increasing-difficulty order (60%) than in increasing-easiness order (40%), $\chi^2(1, N = 200) = 8.00$, $p = .005$. None of the demographic or control variables predicted efficacy ratings.

Suggesting that predictions of efficacy drive people's preferences in this context, in a binomial logistic model, a difference score of efficacy ratings predicted order preference ($\beta = 1.11$, $SE = 0.17$, $p < .001$) even controlling for age, education, gender, and familiarity with the task, none of which predict order preference (βs < 0.36, $SE < 0.45$, $p > .102$).

These Experiment 1a predictors did not endorse the "eat the frog first" mentality, but instead believed that completing tasks in increasing-difficulty (vs. increasing-easiness) order would enhance efficacy and consequently preferred to complete tasks in increasing-difficulty order. Experiment 1b next tests the accuracy of these predictions.

5.3. Exp. 1b method

We preregistered this experiment on the Open Science Framework (https://osf.io/u85fz).

5.3.1. Participants

Given that this study was significantly more expensive than Experiment 1a and we hoped to find at least a medium-sized effect, we decided a priori to recruit a smaller sample size. We preregistered recruiting at least 120 participants in each of three between-subjects conditions. In total, we recruited 363 participants (159 female, $M_{age} = 35.42$ 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. A sensitivity power analysis revealed that we have 80% power to detect an effect size of $d = 0.36$ (at $\alpha = 0.05$) using two-tailed t-tests.

5.3.2. Procedure

We randomly assigned participants (i.e., performers) into one of three experimental conditions (between-subjects): *increasing-difficulty order*, *increasing-easiness order*, and *control*. Performers were told they would be completing three rounds of six analogies each and that they would answer a few questions after each round. They learned exactly the same information described in Experiment 1a and then responded to several attention check items to ensure that they had read and understood the instructions; they had to answer these questions correctly to move on in the study. In order to further incentivize them to try their best in the task, performers earned a $0.02 bonus for each analogy that they correctly answered.

Before beginning the task, performers then saw four easy practice analogies. The first two were fill-in-the-blank (e.g., "helicopter: ____ ::submarine:water" and "____ :eat::tired:sleep."). The second two practice analogies were multiple-choice, intended to get participants used to the format of the rest of the questions. All four practice analogies were very easy so that we could identify performers that did not clearly understand the task. After the multiple-choice analogies, performers immediately received feedback about their response in the form of a green check that said "YES!" (correct answer) or a red X (incorrect answer).

To select the different analogies for each round, we pre-tested 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 Supplementary materials for further details about the pretest. In the *increasing-easiness* condition, performers completed 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 performers completed three rounds of two easy analogies, two medium analogies, and two hard analogies. Within each round, the six analogies were presented in randomized order.

After each round, performers received feedback about how many analogies they answered correctly in that round and in all previous rounds. Overall, performers' total number of correct answers did not meaningfully differ between the *increasing-difficulty* (M = 10.23, 95% CI [9.86, 10.61]), *increasing-easiness* (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 our manipulation did not affect actual skill levels. As designed, and replicating our pretest results, performers 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]), $t(438) = 20.37$, $p < .001$, $d = 1.95$. After each round, performers completed a short survey with a longer post-survey at the end of all three rounds.

5.3.3. Materials (survey)

5.3.3.1. Efficacy experiences. To assess performers' actual efficacy, we asked three questions[3]—“How skilled do you think you 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*]; $\alpha s \geq 0.95$)—at four times throughout the experiment: before the first round and after each round. (Before the first round, the questions changed slightly to the future tense, e.g., “How skilled do you think you will be at these analogy tasks?”).

5.3.3.2. Task and performance beliefs. To determine whether performers believed the tasks were changing in difficulty, we asked, “Did you think the analogies were changing in difficulty?” on a slider scale anchored at 0 (*They were getting way easier*) and 100 (*They were getting way harder*) which started at 50 (*No change in difficulty*). To determine whether they thought their own performance was changing, we asked, “Overall, were you getting better at analogies, getting worse, or staying about the same?” on a slider scale anchored at 0 (*Getting much worse*) and 100 (*Getting better*) which started at 50 (*Staying the same*).

5.3.3.3. Control variables. To control for performers' experience with the task, we 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 enjoyment of task, we also asked, “Overall, how much do you enjoy engaging in these analogy tasks?” (1 = *not at all*, 10 = *very much*). Finally we collected education, income, employment, age, and gender.

5.3.3.4. Attention checks. We 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 s*, *10 s*, *15 s*). Participants had to answer the attention check questions correctly to proceed in the survey.

5.3.3.5. Additional variables. We collected a number of additional exploratory variables (e.g., felt motivation) that are not relevant to the present research and that we report in the Supplementary materials.

5.4. Exp. 1b results

For our analysis strategy, we first removed performers from analyses who clearly did not try to complete the two fill-in-the-blank practice analogies correctly (e.g., writing nonsense words) or who answered both of the extremely easy multiple-choice practice analogies incorrectly. We preregistered this exclusion, but also report the results including all participants in the Supplementary materials. The results do not statistically change. After the exclusion, 327 participants remained (147 female, M_{age} = 35.63 years, 95% CI [34.51, 36.76]). Given that our sample size grew smaller, we re-conducted the power analysis, which showed that we have 80% power to detect an effect size of $d = 0.38$ (at $\alpha = 0.05$) using two-tailed t-tests.

5.4.1. Efficacy experiences

Most critically, and contradicting participants' predictions in Experiment 1a, after completing all three practice rounds performers' efficacy was higher in the *increasing-easiness* condition (M = 6.13, 95% CI [5.70, 6.56]) than in the *increasing-difficulty* condition (M = 4.62, 95% CI [4.15, 5.08]), $t(218) = 4.75$, $p < .001$, $d = 0.64$. Performers' efficacy in the *control* condition (M = 5.26, 95% CI [4.89, 5.64]) fell exactly in between those in the *increasing-easiness* condition, $t(217) = 2.99$, $p = .003$, $d = 0.41$, and those in the *increasing-difficulty* condition, $t(213) = 2.15$, $p = .033$, $d = 0.29$.

5.4.2. Task and performance beliefs

Performers 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 *increasing-easiness* condition recognized that the tasks were getting easier (M = 20.71, 95% CI [15.69, 25.72]), $t(218) = 25.14$, $p < .001$, $d = 3.41$. In the *control* condition, performers thought the tasks were getting slightly harder (M = 64.48, 95% CI [60.93, 68.02]), but not as hard as in the *increasing-easiness* condition, $t(213) = 12.95$, $p < .001$, $d = 1.77$. Performers also reported getting better at analogies in the *increasing-easiness* 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, performers did not think their skill was changing (M = 50.08, 95% CI [46.54, 53.63]), *one-sided* $t(106) = 0.05$, $p = .963$, $d = 0.01$.

5.5. Discussion

Predictors believed that completing tasks in increasing-difficulty order would create more felt efficacy than completing them in increasing-easiness order and thus indicated a preference for starting with their easiest task (Experiment 1a). In fact, performers reported the opposite; completing tasks in increasing-difficulty order (vs. increasing-easiness or random order) reduced efficacy (Experiment 1b). These findings indicate that, just as we hypothesized, people mispredict how the difficulty ordering of tasks influences their efficacy.

However, in order to conclude that people mispredict, predictors must have exactly the same information and expectations as performers. One difference in the paradigm we used in Experiment 1a compared to 1b was that predictors did not actually anticipate completing

[3] Note that the questions measuring efficacy in Experiment 1b are not exactly the same as the questions predicting efficacy in Experiment 1a. The minor difference in wording was an error that we fixed in all other studies, including Experiment S1 and Experiment 4 which replicate the findings reported in Experiments 1a and 1b.

the tasks; thus, their inaccuracy could stem from their predictions being hypothetical. To address this possibility, we ran a follow-up study (Experiment S1) in which participants predicted their feelings of efficacy before completing the tasks in increasing-difficulty and increasing-easiness order and then completed the tasks and subsequently reported their felt efficacy. Thus, the predictors and performers were exactly the same people, with the same expectations about the experiment. Moreover, because predictors knew they would be completing the tasks subsequently, their predictions were potentially more realistic. We observed the same pattern of results in Experiment S1 as in Experiments 1a & 1b, such that participants predicted that increasing-difficulty (vs. increasing-easiness) would enhance their efficacy ($M_{\text{inc-dif}} = 6.32$, 95% CI [6.07, 6.57], $M_{\text{inc-eas}} = 5.89$, 95% CI [5.61, 6.18]), *paired t* (218) = 2.81, $p = .005$, $d = 0.19$, but in reality increasing-easiness enhanced efficacy ($M_{\text{inc-dif}} = 6.32$, 95% CI [5.91, 6.74], $M_{\text{inc-dif}} = 4.69$, 95% CI [4.23, 5.16]), $t(217) = 5.18$, $p < .001$, $d = 0.70$. The expected statistical interaction between predictions and experiences for increasing-difficulty and increasing-easiness order emerged, $F(1, 217) = 40.56$, $p < .001$, $\eta_p^2 = 0.157$. The effect sizes were of similar magnitude as in Experiments 1a ($d = 0.19$) and 1b ($d = 0.64$). This suggests that the potential methodological confound of predictors not actually completing tasks does not meaningfully influence the results. Experiments 2a and 2b thus use the same paradigm as Experiments 1a and 1b.

6. Experiments 2a & 2b: Building self-efficacy in word finds

Experiments 2a and 2b tested the discrepancy between predicted and actual efficacy using a different set of tasks (word finds). Using a different task allows us to test the robustness of the effects observed in Experiments 1a, 1b, and S1. For one, we can examine whether a shorter trajectory (three tasks varying in difficulty-ordering instead of 18) still influences efficacy. We also provided no feedback during task completion, testing whether the effect of task difficulty on efficacy remains without feedback. We expected that our hypothesized effect would emerge despite these changes to the task.

6.1. Exp. 2a method

We preregistered this experiment on the Open Science Framework (https://osf.io/yadks).

6.1.1. Participants

Following our sample size determination in Experiment 1a, we predetermined a sample size of 200 participants. In total, we recruited 202 participants (93 female, $M_{\text{age}} = 35.00$ years, 95% CI [33.41, 36.59]) through Amazon Mechanical Turk who completed a survey in exchange for $0.40. A sensitivity power analysis revealed that we have 80% power to detect an effect size of $d = 0.20$ (at $\alpha = 0.05$) using a two-tailed paired t-test.

6.1.2. Procedure

Predictors in Experiment 2a imagined they were participating in a different MTurk survey in which they would "complete three practice rounds of a word-find task and then compete for an opportunity to win a financial bonus." They were told that a word-find task involves "making as many words as possible out of a string of random letters." For each word-find task, they learned 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. The 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 the letter string was "*XHWYNEAJRTMF*," participants could write words such as: "*near*," "*meat*," "*fret*," "*wart*," "*wharf*," and so on.

We told predictors to imagine completing three practice rounds each composed of a different word-find task which "have been pre-tested [by other people] and are either easy, medium, or hard difficulty." Predictors learned that for the easy word-find practice task, people found an average of 15.1 words, for the medium word-find practice task, people found an average of 10.6 words, and for the hard word-find practice task, people found an average of 8.1 words. The information that we gave predictors about the three word-find tasks was based on a real pre-test that we conducted (see Supplementary materials for details), and they viewed the same word-find tasks that Experiment 2b participants actually did complete.

After learning about the tasks, predictors completed the survey, reporting their preference for task ordering and predicting their efficacy levels after imagining completing the three word-find tasks in increasing-difficulty order and increasing-easiness order (presented in counterbalanced order).

6.1.3. Materials (survey)

6.1.3.1. Efficacy predictions. To collect efficacy predictions, predictors answered three questions: "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*]; $\alpha = 0.93$). We collapsed these three questions into one efficacy index for analysis.

6.1.3.2. Order preferences. To measure predictors' order preference, we 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). We subsequently asked, "Why did you choose [choice]?" (see Table 1 for example reasons).

6.1.3.3. Exploratory and control measures. The survey also contained two exploratory measure and control variables. In the first exploratory measure, we asked predictors to predict how many words they would find in each round. Their predictions did not differ by condition, $p = .566$. In the second exploratory measure, which was collected for a different project, we examined predictors' beliefs about their momentum. These predictions followed the same pattern as their efficacy predictions ($p < .001$). As control variables, we collected predictors' age, education, and gender, and measured their experience with the task: "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).

6.2. Exp. 2a results

Supporting our primary hypothesis and replicating Experiment 1a, predictors 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 increasing-easiness order ($M = 6.46$, 95% CI [6.18, 6.74]), *paired* $t(201) = 4.84$, $p < .001$, $d = 0.34$. Consistently, more participants preferred to complete the tasks in increasing-difficulty order (53%) than in increasing-easiness order (22%), χ^2(1,

N = 152) = 25.29, $p < .001$. 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 binomial logistic model excluding predictors who were indifferent to task ordering, a difference score of efficacy ratings predicted order preference ($\beta = 0.67$, $SE = 0.13$, $p < .001$) even controlling for age, education, gender, and familiarity with the task, none of which predict order preference (βs < 0.39 $SE < 0.43$, $p > .149$). (Note, though, that this was not a preregistered analysis.)

Experiment 2b next tests the validity of these predictions.

6.3. Exp. 2b method

We preregistered this experiment on the Open Science Framework (https://osf.io/2b9fg).

6.3.1. Participants

Given the effects observed in Experiment 1b, we predetermined a sample size of 100 participants for each of two experimental conditions. We recruited 199 participants (84 female, 2 gender non-binary, M_{age} = 36.62 years, 95% CI [35.09, 38.17]) through Amazon Mechanical Turk who completed a survey in exchange for $0.80. A sensitivity power analysis revealed that we have 80% power to detect an effect size of $d = 0.40$ (at $\alpha = 0.05$) using two-tailed t-tests.

6.3.2. Procedure

We randomly assigned participants (i.e., performers) to one of two conditions: *increasing-easiness* or *increasing-difficulty*. Performers learned exactly the same information described in Experiment 2a and then responded to several attention check items to ensure that they had read and understood the rules; they had to answer these questions correctly to move on in the study.

To manipulate the task-difficulty of the letter sets, we first classified the difficulty level of several letter sets that we pre-tested (see Supplementary materials). We selected three letter sets: easy ("AEITFMNLPRYG"; $M_{words\ found}$ = 15.10, 95% CI [13.00, 17.19]), medium ("AEIBTJNCKYDH"; $M_{words\ found}$ = 10.60, 95% CI [9.09, 12.11]), and hard ("EIOBTJNCMYRP"; $M_{words\ found}$ = 8.08, 95% CI [6.76, 9.40]). In the *increasing-easiness* condition, performers completed the hard letter set, then the medium letter set, and then the easy letter set. In the *increasing-difficulty* condition they completed the same letter sets in the opposite order—easy, then medium, then hard. We told performers the difficulty level of each round before they began, just as the Experiment 2a predictors had learned.

To examine actual performance, and hence any changes in performers' real skill levels as they completed the task in different orders, we computed the average scores for each round. Actual practice round scores were not meaningfully different between the *increasing-difficulty* (M = 11.21, 95% CI [10.07, 12.35]) and *increasing-easiness* (M = 11.54, 95% CI [10.54, 12.54]) conditions, $t(166) = 0.43$, $p = .666$, $d = 0.07$, suggesting that the manipulation did not affect actual skill levels. As designed, and replicating our pretest results, performers found more words in the easy round (M = 13.55, 95% CI [12.59, 14.42]) than the medium round (M = 11.36, 95% CI [10.49, 12.22]), $t(334) = 3.55$, $p < .001$, $d = 0.39$, and more words in the medium round than in the hard round (M = 9.23, 95% CI [8.44, 10.01]), $t(334) = 3.60$, $p < .001$, $d = 0.39$. After each round, performers completed a short survey with a longer post-survey at the end of all three rounds.

6.3.3. Material (survey)

6.3.3.1. Efficacy experiences. To assess performers' actual efficacy we asked three questions—"How skilled did you feel at finding words?", "How confident did you feel about finding words?", and "How much do you trust your ability to find words?" (1 = *not at all* [*confident/skilled/*], 10 = *very* [*confident/skilled/much*]; $\alpha s \geq 0.97$)—at four time-points throughout the experiment: before the first round and after each round. (Before the first round, the questions changed slightly to the future tense, e.g., "How skilled do you think you will be at finding words?").

6.3.3.2. Task and performance beliefs. To determine whether performers believed the tasks were changing in difficulty, we asked, "Did you think the letter sets were changing in difficulty?" on a slider scale anchored at 0 (*They were getting way easier*) and 100 (*They were getting way harder*) which started at 50 (*No change in difficulty*). To determine whether they thought their own performance was changing, we asked, "Overall, were you getting better at finding words, getting worse, or staying about the same?" on a slider scale anchored at 0 (*Getting much worse*) and 100 (*Getting better*) which started at 50 (*Staying the same*).

6.3.3.3. Control variables. To control for performers' experience with the task, we 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). To control for enjoyment of task, we also asked, "Overall, how much did you enjoy engaging in these word find tasks?" (1 = *not at all*, 10 = *very much*). Finally, we collected education, income, employment, age, and gender.

6.3.3.4. Attention check questions. We 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 min*, *1.5 min*, *2 min*), "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*). Participants had to answer the attention check questions correctly to proceed in the survey.

6.4. Exp. 2b results

For our analysis strategy, we first removed performers from analyses whose average score across all three rounds was not within one standard deviation below the mean because these performers 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. We preregistered this exclusion, but also report the results including all participants in the Supplementary materials. The results become weaker including all participants. After our preregistered exclusions, 168 participants remained (74 female, 2 gender non-binary, M_{age} = 36.77 years, 95% CI [35.13, 38.41]). Given that our sample size decreased, we re-conducted the sensitivity power analysis, which showed we have 80% power to detect an effect size of $d = 0.43$ (at $\alpha = 0.05$) using two-tailed t-tests.

6.4.1. Efficacy experiences

Contradicting the predictions from Experiment 2a, after completing all three practice rounds, performers' efficacy was marginally higher in the *increasing-easiness* condition (M = 6.36, 95% CI [5.84, 6.88]) than in the *increasing-difficulty* condition (M = 5.64, 95% CI [5.10, 6.18]), $t(166) = 1.91$, $p = .058$, $d = 0.30$. Although this effect of difficulty ordering on efficacy is small, we believe it is robust because three follow-up experiments, which each tested the effect of increasing-difficulty order vs. increasing-easiness order using the same three letter sets (N = 1149), also showed that performers' efficacy was significantly higher in the *increasing-easiness* 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]), $t(1,147) = 12.24$, $p < .001$, $d = 0.72$. We combined the results of the three experiments for analysis but controlled for the experiment number in the regression models using dummy variables. These experiments were specifically designed to examine a broader set of dependent variables and to test whether completing tasks in different difficulty-orders influences performance on a subsequent financially-incentivized competition task of a novel letter set (it does not: difficulty-order condition neither affected performance levels, ($M_{\text{inc-easy}}$ = 26.75, 95% CI [25.82, 27.67], $M_{\text{inc-dif}}$ = 26.17, 95% CI [25.27, 27.08]), $t(1,147) = 0.87$, $p = .384$, $d = 0.05$, nor the actual likelihood of winning the competition, χ^2 (1, N = 1147) < 0.001, $p = .983$). Given that the research question of how efficacy influences future performance is not the focus of the current paper, we simply report the details of each of these three experiments (Experiments S2, S3, and S4) in the Supplementary materials.

6.4.2. Task and performance beliefs

Participants in the *increasing-difficulty* condition recognized that the tasks were getting harder (M = 66.36, 95% CI [62.60, 70.1]), whereas those in the *increasing-easiness* condition recognized that the tasks were getting easier (M = 42.61, 95% CI [37.69, 47.53]), and the two conditions reported significantly different assessments of task-difficulty, $t(166) = 7.63$, $p < .001$, $d = 1.18$. Participants also reported getting better at analogies in the *increasing-easiness* condition (M = 70.15, 95% CI [65.77, 74.54]) more than the *increasing-difficulty* condition (M = 53.82, 95% CI [48.56, 59.08]), $t(166) = 4.74$, $p < .001$, $d = 0.74$.

6.5. Discussion

In conjunction, using controlled and clean experimental designs, Experiments 1 and 2 (and S1) indicate that the difficulty-ordering of task completion influences efficacy quite differently than people believe it will. In addition, we know that participants were making informed predictions because they viewed each task they would complete and received accurate information about the exact success rate for each round of tasks. By explicitly reporting how many words an average participant found in each round, we ensured that the Experiment 2a predictors and Experiment 2b performers had identical knowledge and expectations before completing any tasks. While Experiments 1 and 2 examined efficacy on quantifiable verbal tasks (analogies and word finds), we next turn to the domain of applying for jobs, an important and common task that people complete in their everyday life, in Experiments 3a and 3b.

7. Experiments 3a & 3b: Predicted and actual efficacy when applying to jobs

As with achieving most long-term goals, applying successfully for jobs requires completing tasks that vary in difficulty order (e.g., identifying job opportunities, creating a resume). Experiments 3a and 3b test how people believe that completing such tasks in different orders will influence their efficacy compared to how it actually does.

7.1. Exp. 3a method

7.1.1. Participants

Based on the effect sizes in Experiments 1a and 2a, we predetermined to recruit 100 individuals. In total, 94 undergraduate students (68 female, 3 non-binary, M_{age} = 18.74 years, 95% CI [18.60, 18.89]) from a psychology laboratory pool at a public west coast university participated in exchange for an entry to a lottery for a $50 gift card. A sensitivity power analysis revealed that we have 80% power to detect an effect size of $d = 0.29$ (at $\alpha = 0.05$) using a two-tailed paired t-test.

7.1.2. Procedure

We asked participants in Experiment 3a (i.e., predictors) to "imagine that you are applying for a job" and to "review and evaluate nine tasks that are often required to complete a job application." Before making any predictions, participants first read a detailed description of each task (in randomized order) and rated them on four 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 min and 60 min with anchors every 10 min to indicate how long they thought the task would take to complete. We collected these ratings to better understand how the tasks varied not only on perceived difficulty but also other aspects.

The nine tasks were: 1) to complete a job application form with some basic identifying information (e.g., first name, last name, email address, and so on); 2) to create profiles of their three most recent jobs (entering: employer, position, start date, end date, and website); 3) to identify and provide contact information for three references and to draft the text of an email to send to each person asking for him/her to act as a character reference for them; 4) to write a short cover letter (with template and example provided); 5) to indicate availability for a phone screen with a recruiter by checking their calendar and marking the available times; 6) to describe a challenge they faced or a mistake they made in a previous job and what they learned from the experience (in order to prepare for interviews); 7) to create an objective sentence for their resume (with template and examples provided); 8) to provide a list of five duties they had executed in prior jobs in bullet point format (with template and examples provided); 9) to describe any clubs and other extracurricular activities they participated in during college or high school and explain their involvement (in order to prepare for interviews). For complete descriptions, see the Supplementary materials.

After viewing and rating all nine tasks, predictors then ranked the tasks from easiest to hardest and completed the prediction survey, reporting their preference for task ordering and predicting their efficacy levels after imagining completing the tasks in increasing-difficulty order and increasing-easiness order (counterbalanced).

7.1.3. Materials (survey)

7.1.3.1. Efficacy predictions. Predictors imagined that they were "assigned to see the tasks in the following order: Easy to hard [Hard to easy]" and answered three questions: "If you were assigned to complete the tasks from easy to hard [hard to easy], how skilled do you think you would feel about applying for jobs after completing all the tasks?", "If you were assigned to complete the tasks from easy to hard [hard to easy], how confident do you think you would feel about applying for jobs after completing all the tasks?", "If you were assigned to complete the tasks from easy to hard [hard to easy], how much would you trust your ability to apply to jobs after completing all the tasks?" (1 = *not at all* [*skilled/confident*], 10 = *very skilled* [*confident/much*]; $\alpha = 0.91$). We collapsed these three questions into one efficacy index for analysis.

7.1.3.2. Order preferences. To measure predictors' order preference, we asked: "If you were forced to choose to complete all nine tasks in either decreasing difficulty order (starting with the hardest task and ending with the easiest) or increasing difficulty order (starting with the easiest and ending with the hardest), which would you choose?" (*Easy to hard*; *Hard to easy*). For exploratory analysis we also asked "Why did you choose [choice] as your preferred order?" (see Table 1 for example reasons).

7.1.3.3. Exploratory and control measures. As an exploratory measure, in each condition we also asked, "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?" (1 = *not at all likely*, 10 = *very likely*). Predictors 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 reported their class year, major, whether they were currently applying for jobs, age, and gender.

7.2. Exp. 3a results

Yet again, predictors 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 increasing-easiness order ($M = 6.02$, 95% CI [5.67, 6.37]), *paired* $t(93) = 3.02$, $p = .003$, $d = 0.31$. As in Experiments 1a and 2a, significantly more predictors preferred to complete the tasks in increasing-difficulty order (81%) than increasing-easiness order (19%), χ^2 (1, $N = 94$) = 35.79, $p < .001$. Using a difference score, we also find that expected efficacy predicted preferences in a binomial logistic regression ($\beta = 0.97$, $SE = 0.30$, $p = .001$) even controlling for age, gender, year in school, familiarity with applying for jobs, and whether they are currently applying for a job. (The first four control variables did not predict order preferences at all, $\beta s < 0.68$, $SE < 1.04$, $p > .421$, whereas current job application status marginally did, $\beta = -2.75$, $SE = 1.48$, $p = .063$.)

Examining our exploratory measure, 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 increasing-easiness order ($M = 6.34$, 95% CI [5.88, 6.80]), $t(401) = 2.81$, $p = .005$, $d = 0.41$, suggesting that they believed their greater efficacy would translate into action.

Experiment 3b next examines the actual consequences of completing the tasks in differing orders.

7.3. Exp. 3b method

We preregistered this experiment on the Open Science Framework (https://osf.io/am36v).

7.3.1. Participants

Following our sample size determination in Experiment 2b, we predetermined to recruit 100 participants in each of two between-subjects conditions. In total, 200 undergraduate students (150 female, 2 non-binary, $M_{age} = 20.96$ years, 95% CI [20.78, 21.13]) from a psychology laboratory pool at a public west coast university participated in exchange for $15. A sensitivity power analysis revealed that we have 80% power to detect an effect size of $d = 0.40$ (at $\alpha = 0.05$) using two-tailed t-tests.

7.3.2. Procedure

From the set of nine tasks tested in Experiment 3a, we selected five tasks that sufficiently varied in difficulty ratings but not on other metrics (i.e., pleasantness, importance, and desire to complete; see Table 2). The tasks were: indicating times the participant would be available for a phone interview (Task 1), filling out a form with information about three previous jobs (Task 2), detailing three previous job duties (Task 3), creating an objective sentence for a resume (Task 4), and writing about a challenge participants had faced in the past (Task 5). Specifically, all of the tasks varied in rated difficulty level except for Tasks 3 and 4 (which were the same difficulty level), $ts > 3.10$, $ps < 0.002$, $ds > 0.46$, but the tasks did not vary in perceived pleasantness, $ts > 1.85$, $ps > 0.066$, $ds < 0.27$, nor in perceived importance, $ts < 1.90$, $ps > 0.059$, $ds < 0.28$ (except for Task 4, which was considered less important than the others, $ts > 2.10$, $ps < 0.038$, $ds > 0.31$), nor in wanting to complete, $ts < 0.88$, $ps > 0.380$, $ds < 0.13$ (except for Task 1, which participants wanted to complete more than the other tasks, $ts > 2.58$, $ps < 0.011$, $ds > 0.38$). This process ensured that we manipulated the difficulty ordering of tasks without manipulating these other aspects of the tasks.

In order to determine the ordering of tasks that the Experiment 3b performers would complete, we first showed performers all five tasks in randomized order and asked the performers to a) rate them on difficulty, pleasantness, importance, and desire to complete them and to b) rank them from easiest to hardest (see survey details below). We randomly assigned performers into either the *increasing-difficulty* or the *increasing-easiness* condition, ordering the tasks based on performers' own difficulty rankings.

After completing each task, performers answered a short survey. After completing all five tasks, performers answered a longer post-survey. The tasks and surveys were administered in the laboratory to ensure that performers were engaged the entire time.

7.3.3. Materials (survey)

7.3.3.1. Task ratings and rankings. Prior to completing the tasks, performers rated them on four measures: "How difficult this task is", "How pleasant this task is", "How important this task is in the job application process", and "How much you want to complete this task" (1 = *not at all*, 10 = *very*). We subsequently asked performers to rank the difficulty order of the tasks: "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)."

7.3.3.2. Order preferences. Unlike in Experiments 1b and 2b, here we collected performers' preferences 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 asked them to report why they made that choice.

7.3.3.3. Efficacy experiences. To assess performers' actual efficacy, we asked two different sets of efficacy questions (three questions each), 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, we 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*]; $\alpha s \geq 0.95$).

Table 2
Means and 95% Confidence Intervals for Variables of Interest for Chosen Tasks in Experiment 3b.

	Task 1	Task 2	Task 3	Task 4	Task 5
Difficulty	2.20 (1.89, 2.52)	2.98 (2.59, 3.36)	4.90 (4.49, 5.32)	4.90 (4.53, 5.28)	6.02 (5.56, 6.48)
Pleasantness	4.74 (4.32, 5.17)	4.24 (3.82, 4.66)	4.57 (4.21, 4.94)	4.80 (4.38, 5.22)	4.31 (3.86, 4.76)
Importance	7.63 (7.13, 8.13)	7.01 (6.61, 7.41)	7.35 (6.97, 7.74)	6.36 (5.90, 6.83)	7.23 (6.80, 7.67)
Want to complete	6.14 (5.62, 6.66)	5.14 (4.68, 5.59)	4.94 (4.50, 5.38)	5.22 (4.75, 5.70)	5.04 (4.55, 5.53)

7.3.3.4. Exploratory motivation measures. We collected several exploratory measures of motivation: "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*). None of these measures varied by experimental condition and are not discussed further in the main text; see the Supplementary materials for analyses.

7.3.3.5. Task beliefs. We asked performers, "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*).

7.3.3.6. Control variables. We included the following control variables at the end of the survey: "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*). We also collected two measures 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, we collected participants' years in school, majors, ages, and genders.

7.4. Experiment 3b results

7.4.1. Task ratings and rankings

Prior to completing the tasks, performers' ratings looked similar to the Experiment 3a predictors' ratings (see the Supplementary materials for details). But after completing the tasks, performers' overall task difficulty ratings did not differ by condition ($M_{\text{inc-easy}}$ = 3.89, 95% CI [3.64, 4.13], $M_{\text{inc-dif}}$ = 3.99, 95% CI [3.71, 4.26]), $t(198) = 0.52$, $p = .602$, $d = 0.07$, indicating that participants had similar experiences on this metric in aggregate.

7.4.2. Order preference

Consistent with the results in Experiments 1a–3a, performers preferred to order the tasks in increasing-difficulty order: 65.5% of performers selected increasing-difficulty order and the remaining 34.5% selected increasing-easiness order, $\chi^2(1, N = 200) = 19.22, p < .001$.

7.4.3. Efficacy experiences

Unlike in other experiments, performers' efficacy for applying to jobs did not differ by condition ($M_{\text{inc-easy}}$ = 6.23, 95% CI [5.84, 6.62], $M_{\text{inc-dif}}$ = 6.04, 95% CI [5.64, 6.43]), $t(198) = 0.67$, $p = .506$, $d = 0.10$, nor did their efficacy for getting an interview ($M_{\text{inc-easy}}$ = 6.05, 95% CI [5.64, 6.47], $M_{\text{inc-dif}}$ = 6.08, 95% CI [5.63, 6.53]), $t(198) = 0.09$, $p = .926$, $d = 0.01$. Efficacy ratings were also not moderated by performers' stated ordering preference and there was no interaction between order preference and condition, $F\text{s} < 0.44$, $p\text{s} > 0.508$, $\eta^2\text{s} < 0.07$.

7.4.4. Performance

Two coders blind to hypothesis rated the quality of the completed tasks (1 = *extremely low quality*, 10 = *extremely high quality;* $\alpha = 0.65$); there was no difference in performance by condition when analyzing either coders' ratings or when combining the two ratings, $t\text{s} < 0.70$, $p\text{s} > 0.487$, $d\text{s} < 0.10$.

7.4.5. Task beliefs

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 *increasing-easiness* 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, $t(198) = 11.55$, $p < .001$, $d = 1.64$.

7.5. Discussion

Overall, differences in the task difficulty-ordering did not causally influence task-relevant efficacy in Experiment 3b, despite participants' predictions that they would. There are many possible reasons for why task difficulty-ordering causally influenced efficacy ratings in Experiments 1b and 2b but in Experiment 3b. For instance, word finds and analogies allow for easily quantifiable metrics of performance and success, while applying for jobs has only a single success metric (receiving an offer). Similarly, perhaps task-ordering primarily changes efficacy with more immediate feedback than is available in the job application paradigm. We consider further possibilities in the General Discussion.

Regardless of the reason why the task-difficulty order showed no effect on self-efficacy in Experiment 3b, the gap between predictions and reality again indicates that individuals misunderstand how task order difficulty influences their efficacy levels. We conclude that even for common and consequential tasks like applying for jobs, individuals misunderstand how the order in which they complete tasks of varying difficulty levels can influence their efficacy.

8. Experiment 4: Enhancing the accuracy of mental simulations

Experiment 4 tests our hypothesis that one reason why people mispredict how task difficulty-order influences their efficacy levels is because they fail to fully simulate their trajectory of efficacy experiences, imagining the beginning of the trajectory more than the end. We not only test how people naturally tend to make predictions about their efficacy but also examine a potential intervention to get them to think more about how the end of the task trajectory, not just the beginning, influences their self-efficacy.

8.1. Method

We preregistered this experiment on the Open Science Framework (https://osf.io/hx6qz).

8.1.1. Participants

Following our sample size determination in Experiments 3a and 3b, we predetermined a sample size of 100 participants in each of five experimental conditions. In total, we 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. A sensitivity power analysis revealed that we have 80% power to detect an effect size of $d = 0.40$ (at $\alpha = 0.05$) using two-tailed t-tests.

8.1.2. Procedure

We adapted the paradigm from Experiments 1a and 1b, using the analogy tasks described in those experiments. We randomly assigned participants to one of five between-subject experimental conditions: *low-simulation prediction*, *moderate-simulation prediction*, *high-simulation prediction*, *increasing-difficulty experience*, or *increasing-easiness experience*. Each of the three prediction conditions also contained two within-subjects conditions in counterbalanced order: *increasing-difficulty* and *increasing-easiness*. The two *experience* conditions followed the same procedure described in the Experiment 1b Method.

The *low-simulation prediction* condition followed the same procedure described in Experiment 1a Method, whereby predictors first saw the analogies in each round in random order and subsequently predicted how much efficacy they would feel if they completed the rounds in increasing-difficulty order and in increasing-easiness order (presented in counterbalanced order). To ensure participants viewed the analogy rounds carefully, we presented each round for 10 s.

The *moderate-simulation prediction* condition nudged participants to simulate the experience conditions more fully by first showing the predictors the analogies for each round in the same order for which they were making a prediction, and subsequently collecting their prediction. For instance, participants viewed the easy round, medium round, and hard round (for 10 s each) and then predicted their increasing-difficulty efficacy; they then viewed the same analogies in the opposite order and predicted their increasing-easiness efficacy.

The *high-simulation prediction* condition was intended to get participants to focus most on the end of the trajectory instead of just the beginning of the trajectory. In addition to viewing each round in the same order for which they were predicting, predictors in the *high-simulation prediction* condition predicted their efficacy after seeing each round of analogies, thus forcing them to not just view but also make predictions regarding the entire trajectory.

8.1.3. Materials (survey)

8.1.3.1. Efficacy experiences and predictions. To collect experienced and predicted efficacy, participants answered three questions, modified slightly for each condition. All conditions ended with the same three questions, modified for tense in the prediction [experience] conditions: "…how skilled would [do] you think you are at answering analogies correctly?", "…how confident would [do] you feel about answering analogies correctly?", "…how much would [do] you trust your ability to answer analogies correctly?" (1 = *not at all* [*skilled/confident/*], 10 = *very* [*skilled/confident/much*]; $\alpha s > 0.92$). In the *low-simulation prediction* condition, the questions began "If you were assigned to complete the rounds from easy to medium to hard [hard to medium to easy]…" In the *moderate-simulation prediction* condition, we asked "If you were assigned to complete the rounds in the order you just saw (easy to medium to hard) [(hard to medium to easy)]…" In the *high-simulation prediction* condition we asked about efficacy three times. After the first round, we asked "After first completing the easy [hard] round…", after the second "After next completing the medium round…", and finally "After having completed the easy [hard] then medium then hard [easy] rounds…" In the *experience* conditions, the questions began "After completing the first round…", "After completing the second round…", and "After completing all three rounds…".

8.1.3.2. Task and performance beliefs. To determine whether performers believed the tasks were changing in difficulty, we asked, "Did you think the analogies were changing in difficulty" on a slider scale anchored at 0 (*They were getting way easier*) and 100 (*They were getting way harder*) which started at 50 (*No change in difficulty*). To determine whether they thought their own performance was changing, we asked, "Overall, were you getting better at analogies, getting worse, or staying about the same?" on a slider scale anchored at 0 (*Getting much worse*) and 100 (*Getting better*) which started at 50 (*Staying the same*).

8.1.3.3. Control variables. For all participants, we 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*), and demographics (education, income, employment, age, and gender).

8.2. Results

We preregistered removing participants who performed poorly in four practice analogies. We excluded sixteen participants by these criteria. Including these participants does not change the significance of the results.

8.2.1. Efficacy experiences

Replicating the findings from Experiment 1b and 2b, participants in the *increasing-easiness* condition ($M = 6.42$, 95% CI [5.94, 6.90]) reported higher felt efficacy than those in the *increasing-difficulty* condition ($M = 4.47$, 95% CI [3.98, 4.97]), $t(182) = 5.59$, $p < .001$, $d = 0.83$.

8.2.2. Efficacy predictions

Replicating the findings from Experiments 1a, 2a, and 3a, participants in the *low-simulation prediction* condition predicted that completing tasks in increasing-difficulty order ($M = 6.47$, 95% CI [6.16, 6.78]) would lead to greater efficacy than completing them in increasing-easiness order ($M = 4.88$, 95% CI [4.47, 5.29]), $t(198) = 6.17$, $p < .001$, $d = 0.88$. Participants in the *moderate-simulation prediction* condition predicted no difference in feelings of efficacy for the increasing-easiness order ($M = 5.62$, 95% CI [5.19, 6.06]) or increasing-difficulty order ($M = 6.01$, 95% CI [5.60, 6.42]), $t(196) = 1.29$, $p = .200$, $d = 0.18$. Finally, participants in the *high-simulation prediction* condition correctly predicted that completing these analogy tasks in increasing-easiness order ($M = 7.79$, 95% CI [7.42, 8.17]) would lead to greater feelings of efficacy than completing them in increasing-difficulty order ($M = 4.57$, 95% CI [4.10, 5.05]), $t(204) = -10.52$, $p < .001$, $d = -1.48$, supporting our hypothesis that making people focus on the end as well as the beginning of the trajectory would change their expected efficacy. See Fig. 1.

To compare the three prediction conditions to one another, we created a difference score by subtracting the increasing-difficulty prediction from the increasing-easiness prediction. As hypothesized, there was a difference by condition, $F(2, 299) = 146.4$, $p < .001$, $\eta^2 = 0.49$, such that participants in the *high-simulation prediction* condition ($M = 3.22$, 95% CI [2.68, 3.76]) showed greater difference between efficacy predictions than those in the *moderate-simulation prediction* condition ($M = -0.39$, 95% CI [-0.69, -0.09]), $t(200) = 11.45$, $p < .001$, $d = 1.62$, which in turn showed greater difference than those in the *low-simulation prediction* condition ($M = -1.59$, 95% CI [-1.94, -1.24]), $t(197) = 5.20$, $p < .001$, $d = 0.74$.

8.2.3. Task and performance beliefs

In the experience condition, we 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 ($M = 90.33$, 95% CI [87.38, 93.27]), whereas those in the *increasing-easiness* condition recognized that the tasks were getting easier ($M = 25.72$, 95% CI [19.15, 32.30]). Participants also reported getting better at analogies in the *increasing-easiness* condition ($M = 70.02$, 95% CI [66.08, 73.96]) and getting worse at analogies in the *increasing-difficulty* condition ($M = 30.28$, 95% CI [25.34, 35.22]), $t(196) = 12.43$, $p < .001$, $d = 1.78$, even though their overall scores across the three practice rounds were no different ($M_{\text{inc-easy}} = 10.25$, 95% CI [9.83, 10.67], $M_{\text{inc-dif}} = 10.38$, 95% CI [10.00, 10.75]), $t(196) = 0.46$, $p = .647$, $d = 0.07$.

8.3. Discussion

This experiment provides evidence that failing to fully simulate the trajectory of task-difficulty—specifically, focusing more on the beginning of the sequence than the end of the sequence—can help explain why people mispredict how task ordering influences their efficacy. Although individuals always had the opportunity to simulate their

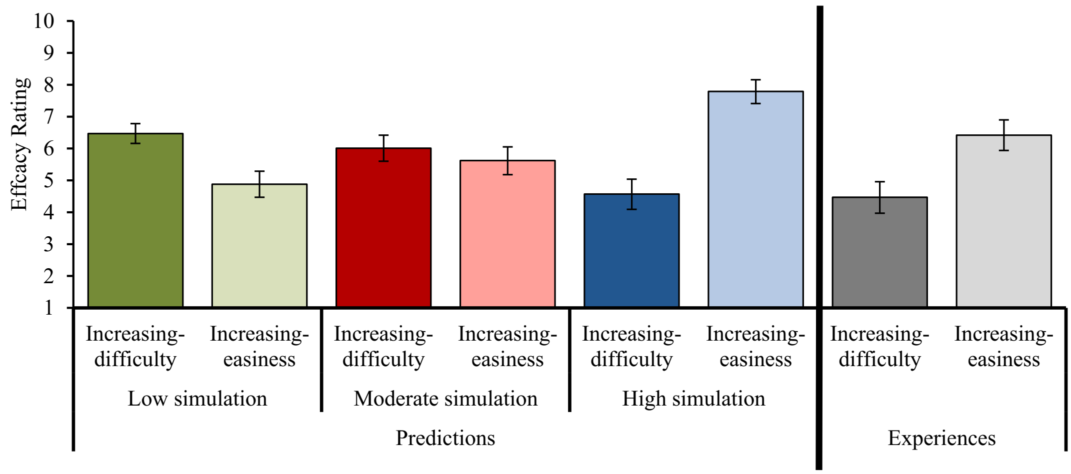

Fig. 1. Predicted and experienced efficacy ratings by condition in Experiment 4. Error bars represent 95% confidence intervals around the mean.

efficacy in each task round to form predictions about the efficacy trajectory, this experiment suggests that people do not naturally do so. Instead, it appears that they consider the beginning of the sequence when predicting their resulting efficacy, failing to think enough about how the end of the sequence will also impact efficacy. Thus, when we made predictors actually report their expected efficacy after each round (not only having them view the task but also think about their resulting efficacy), they correctly predicted that completing tasks in increasing-easiness order leads to higher felt efficacy than doing so in increasing-difficulty order.

9. General discussion

Folk wisdom suggests that, to improve self-efficacy, people should “eat the frog” (do the most difficult thing) first, yet people often prefer to delay difficult tasks. The present research shows that people do not see the value in “eating the frog” first—they instead believe that completing tasks in increasing-difficulty order (eating the frog last) will enhance their self-efficacy. But in reality, completing tasks in increasing-difficulty order harms efficacy more than helps it. Our data suggest that, to maximize efficacy, people should start with their hardest task first; as their task load becomes increasingly easier for them to complete, their efficacy will likewise grow. We further suggest a reason for why people misunderstand how their efficacy changes over time: they simulate the beginning of the task-difficulty sequence more than the end. When nudged to simulate task sequences more fully, such that the end is more apparent, people become better at predicting their actual efficacy levels.

9.1. Theoretical contribution

This work contributes to the existing literature in a number of ways. First, we examine a new predictor of efficacy, task-ordering. Whereas prior research has focused on Bandura's (1977) four proposed sources of self-efficacy (past performance, modeled behavior, verbal persuasion, and physiological reactions), identifying predictors such as goal-setting (Locke and Latham, 2002), others' expectations (Rosenthal and Jacobson, 1968), and coaching (Baron and Morin, 2010), we instead consider the phenomenological experience of how sets of tasks are completed, keeping the overall performance level the same. To our knowledge, this is the first empirical evidence that simply reversing the order in which one completes tasks can causally influence feelings of efficacy.

Secondly, our findings shed light on both procrastination and precrastination. While procrastination has been shown to potentially stem from many sources such as task characteristics (e.g. aversiveness), individual differences (e.g. impulsiveness, achievement-motivation, self-control), or mood regulation, to name a few (Steel, 2007), our results suggest another reason: that people simply do not see much benefit in completing their hardest task first. Getting people to look beyond the initial pain of completing the first task to see the psychological value (e.g., garnering efficacy) of completing their easier tasks later may help to diminish procrastination.

Although the majority of the participants in our studies did not prefer to complete the hardest task first (more consistent with *pro*crastination than *pre*crastination), a sizable minority (30.5%) did report preferring increasing-easiness order. We asked participants to report why they made their choice in all studies (Table 1), providing qualitative insights into the possible reasons why some individuals preferred increasing-easiness order. The participants who preferred increasing-easiness order provided explanations such as, “I like to get the hard work out of the way” and “I would rather do the easier stuff at the end because I know it will require less energy/thinking,” which are consistent with precrastination findings (Rosenbaum et al., 2014; VonderHaar et al., 2019).

Finally, we identify a potential 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 choices that do not maximize happiness in contexts such as taking a new job or moving homes (Hsee and Zhang, 2004), seeking revenge (Carlsmith et al., 2008), or taking pre-emptive actions to avoid regret (Gilbert et al., 2004). Our findings suggest that people may be able to more accurately predict their future psychological states if they focus more on simulating the entire trajectory (the end as well as the beginning) instead of just the start of the trajectory.

9.2. Limitations and future directions

Beyond providing new theoretical insights, this research also leaves unanswered questions and opportunities for future work. First, although not the focus of our primary research questions, in a number of studies we collected participants' performance on their tasks (i.e., correct analogy questions and number of words found) as well as their efficacy. Not only did task performance predict participants' reported efficacy immediately after completing that task, but participants' efficacy after completing one task positively predicted their performance on the following task (controlling for the condition and experiment number with dummy-coded variables, β = 0.07, 95% CI [0.00, 0.14], p = .041). Although the association between efficacy and performance in the current studies is highly confounded by the experimental manipulations (the tasks getting easier or harder), the broader question of when and how self-efficacy—and psychological momentum—influences future performance is important. Indeed, this question has been debated for decades.

A number of theories propose that felt efficacy and/or momentum should influence performance outcomes in domains that require both skill and effort, similar to those we tested (Adler, 1981; Briki et al., 2012, 2014; Briki and Markman, 2018; Iso-Ahola and Dotson, 2014, 2016; Markman and Guenther, 2007). Empirical demonstrations of this proposal (i.e., the "hot hand effect") come from sports domains such as racquetball (Iso-Ahola and Mobily, 1980), tennis (Jackson and Mosurski, 1997; Ransom and Weinberg, 1985; Silva III et al., 1988; Weinberg et al., 1981), volleyball (Raab et al., 2012; but see Miller and Weinberg, 1991), bowling (Dorsey-Palmateer and Smith, 2004; Yaari and David, 2012), and basketball (Forthofer, 1991; Larkey et al., 1989; Mace et al., 1992; Yaari and Eisenmann, 2011), and are robust even when controlling for players' actual skill levels. However, some studies also report the opposite effect (for review, see Habbert, 2019; Habbert et al., 2020). For example, batting averages in baseball tend to regress toward players' means (Albright, 1993; Schall and Smith, 2000), and some experiments that manipulated momentum found that positive momentum reduced effort (e.g., among cyclists, Perreault et al., 1998; and rowers, Den Hartigh et al., 2014). Future work should more deeply investigate this empirical puzzle, which could speak to whether task difficulty order could influence not only felt efficacy but also actual performance in the future.

Second, although we observed only one null effect of task difficulty ordering on efficacy (for job application tasks in Experiment 3b), this null effect suggests several possible moderators that could exist, which future research could test. One moderator could be the task type; for novel tasks, people may need to start with the easiest task in order to learn how to complete it before tackling the hardest one. Another moderator could be task feedback. A key difference between the analogy and word find tasks (Experiments 1 and 2) versus the job application tasks (Experiment 3) was the presence of feedback—whereas performers received feedback on the former tasks, they received none on the latter task. It may be not only the subjective feeling that tasks are becoming easier, but specifically experiencing success in task completion (which requires feedback) that elicits efficacy. We think it could be worth studying whether task difficulty-ordering still influences efficacy among more day-to-day tasks that lack correct answers or clear feedback.

Another direction for future research is to examine how self-efficacy influences goal-setting and motivation. Might people benefit from setting more challenging goals at first and easier goals later in order to establish and build efficacy? This possibility would diverge from predictions made by goal-setting theories, which advocate consistently setting challenging goals to improve performance (Locke and Latham, 2002). Other research has found that competitors tend to slack off when they feel like they are far enough ahead or have accomplished enough (Carver and Scheier, 2012; Koo and Fishbach, 2014; Louro et al., 2007). This might suggest a possible cost to completing tasks in increasing-easiness order because people could feel as though they have accomplished enough after finishing their hardest tasks first and not bother to complete the easier ones later. Future research could integrate the current paper's task-ordering findings with goal-setting and motivation theories in order to better understand how people experience progress versus setbacks when they are striving to attain their goals (e.g., Fishbach et al., 2003; Fishbach and Shah, 2006; Koo and Fishbach, 2014; Schroeder and Fishbach, 2015).

Other potential costs of completing tasks in increasing-easiness order could also be worth investigating, providing a counterpoint to the benefits identified in the current paper (i.e., improved efficacy). For example, studies by Weinstein and Roediger (2010, 2012) showed that participants were less optimistic and felt worse about their *past* performance when they had previously completed the tasks in increasing-easiness (vs. increasing-difficulty) order, suggesting one may have more negative memories when completing tasks in increasing-easiness order. Alternatively, if the first task is too hard, failure could harm efficacy enough that people choose not to attempt the easier tasks.

9.3. Conclusion

Building efficacy is a fundamental human need. Here we show that a basic component of people's everyday lives—the way they order skill-based tasks—influences their efficacy in ways that they may not expect. This research provides evidence that "eating the frog" (doing the hardest task) first, although intuitively unpalatable, can ultimately enhance efficacy. People's tendency to delay the difficult incurs unexpected costs to self-worth.

Open practices

All data, code, and survey materials are available in the Open Science Framework (OSF) repository for this project (https://osf.io/puxfq).

Acknowledgments

We thank Loryn Kanemaru and Trevor Wesolowski for their support, and Don Moore, Cameron Anderson, and Dacher Keltner for their helpful comments on the paper. We also thank Jennifer Abel, Lila Stoddard, Stephanie Millot, and Ashlyn Wright for their assistance conducting the studies. The research was financially supported by the University of California, Berkeley Haas School of Business.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2020.104032.

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.
Attali, Y. (2013). Perceived hotness affects behavior of basketball players and coaches. *Psychological Science, 24*(7), 1151–1156.
Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
Bandura, A. (1988). Self-efficacy conception of anxiety. *Anxiety Research, 1*(2), 77–98.
Baron, L., & Morin, L. (2010). The impact of executive coaching on self-efficacy related to management soft-skills. *Leadership and Organization Development Journal, 31*(1), 18–38.
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. *International 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.
Briki, W., & Markman, K. D. (2018). Psychological momentum: The phenomenology of goal pursuit. *Social and Personality Psychology Compass, 12*(9), 1–14.
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. (2012). Cybernetic control processes and the self-regulation of behavior. *The Oxford Handbook of Human Motivation* (pp. 28–42). .
Cornelius, A., Silva, 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.
Deci, E. L., & Ryan, R. M. (1985). *Self-determination and intrinsic motivation in human behavior.* New York: Plenum.
Den Hartigh, R., Van Der Sluis, J., Zaal, F., & Gernigon, C. (2017). Psychological momentum shapes athletes' affordances. *Paper presented at congress of the French Society of Sport Psychology (SFPS)*. Montpellier: France.
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.
Den Hartigh, R. J., Van Geert, P. L., Van Yperen, N. W., Cox, R. F., & Gernigon, C. (2016). Psychological momentum during and across sports matches: Evidence for interconnected time scales. *Journal of Sport and Exercise Psychology, 38*(1), 82–92.
Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *The American Statistician, 58*(1), 38–45.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Feltz, D. L., & Weiss, M. R. (1982). Developing self-efficacy through sport. *Journal of Physical Education, Recreation & Dance, 53*(3), 24–36.

Fishbach, A., Friedman, R., & Kruglanski, A. (2003). Leading us not unto temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology, 84*(2), 296–309.

Fishbach, A., & Shah, J. Y. (2006). Self-control in action: Implicit dispositions toward goals and away from temptations. *Journal of Personality and Social Psychology, 90*(5), 820–832.

Forthofer, R. (1991). Streak shooter—The sequel. *Chance, 4*(2), 46–48.

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

Habbert, R. (2019). *Miscalibration, missing attributions, and mispredictions: An exploration of momentum, efficacy, and performance expectations.* UC Berkeley. ProQuest ID: Habbert_berkeley_0028E_18777. Merritt ID: ark:/13030/m5dz5c97. Retrieved from https://escholarship.org/uc/item/1t21z01g.

Habbert, R., Schroeder, J., & Markman, K. (2020). The effect of psychological momentum on performance expectations versus performance outcomes. Retrieved from psyarxiv.com/nzfua.

Harris, C. R. (2012). Feelings of dread and intertemporal choice. *Journal of Behavioral Decision Making, 25*(1), 13–28.

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.

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.

Kahneman, D., & Snell, J. S. (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.

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.

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.

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.

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

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.

Mowrer, O. (1960). *Learning theory and behavior.* John Wiley Sons Inc.

O'Donoghue, T., & Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics, 116*(1), 121–160.

Perreault, S., Vallerand, R. J., Montgomery, D., & Provencher, P. (1998). Coming from behind: On the effect of psychological momentum on sport performance. *Journal of Sport & Exercise Psychology, 20*(4), 421–436.

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.

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.

Rosenbaum, D. A., Gong, L., & Potts, C. A. (2014). Pre-crastination: Hastening subgoal completion at the expense of extra physical effort. *Psychological Science, 25*(7), 1487–1496.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3*(1), 16–20.

Ryan, R. M., & Deci, E. L. (2000). The darker and brighter sides of human existence: Basic psychological needs as a unifying concept. *Psychological Inquiry, 11*(4), 319–338.

Schall, T., & Smith, G. (2000). Do baseball players regress toward the mean? *The American Statistician, 54*(4), 231–235.

Schroeder, J., & Fishbach, A. (2015). How to motivate yourself and others? Intended and unintended consequences. *Research in Organizational Behavior, 35*, 123–141.

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.

Silva, J. M., III, Hardy, C. J., & Crace, R. K. (1988). Analysis of psychological momentum in intercollegiate tennis. *Journal of Sport and Exercise Psychology, 10*(3), 346–354.

Solomon, L. J., & Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology, 31*(4), 503–509.

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.

Tracy, B. (2007). *Eat That Frog!: 21 Great Ways to Stop Procrastinating and Get More Done in Less Time.* Berrett-Koehler Publishers.

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.

VonderHaar, R. L., McBride, D. M., & Rosenbaum, D. A. (2019). Task order choices in cognitive and perceptual-motor tasks: The cognitive-load-reduction (CLEAR) hypothesis. *Attention, Perception, & Psychophysics, 81*(7), 2517–2525.

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.

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.