Check for updates

*Research Article*

Psychological Science
2017, Vol. 28(7) 1000–1015
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617701749
www.psychologicalscience.org/PS

# Worth the Wait? Leisure Can Be Just as Enjoyable With Work Left Undone

**Ed O'Brien and Ellen Roney**

## Abstract

Four studies reveal that (a) people hold a robust intuition about the order of work and leisure and that (b) this intuition is sometimes mistaken. People prefer saving leisure for last, believing they would otherwise be distracted by looming work (Study 1). In controlled experiments, however, although subjects thought their enjoyment would be spoiled when they played a game before rather than after a laborious problem-solving task, got a massage before rather than after midterms, and consumed snacks and watched videos before rather than after a stressful performance, in reality these experiences were similarly enjoyable regardless of order (Studies 2 through 4). This misprediction was indeed mediated by anticipated distraction and was therefore attenuated after people were reminded of the absorbing nature of enjoyable activities (Studies 3 and 4). These studies highlight the power of hedonic experience within the moment of consumption, which has implications for managing (or mismanaging) everyday work and leisure. People might postpone leisure and overwork for future rewards that could be just as pleasurable in the present.

## Keywords

work, leisure, enjoyment, intuition, hedonic experience, open data, open materials

Received 10/5/16; Revision accepted 3/6/17

For many people, everyday life is a balancing act between work and leisure. Suppose that by the end of the day you plan to finish grading as well as finish a movie. How would you allocate your time? Which of these activities would you complete first?

Previous research suggests that most people would hope to complete work tasks (e.g., the grading) before, rather than after, engaging in leisure (e.g., the movie). People prefer getting negative events over with quickly and saving positive events for the end of a series (Loewenstein & Prelec, 1993; Novemsky & Ratner, 2003; O'Brien & Ellsworth, 2012; Ross & Simonson, 1991). People also value deservingness (Feather, 1999) and search for cues of work and effort to justify past indulgences (Khan & Dhar, 2006; Kivetz & Simonson, 2002; Xu & Schwarz, 2009). These findings suggest that, on average, people may hold a clear intuitive preference for activity order: work first, leisure second.

In the current studies, we sought to document this intuition and test whether it pays off. On the one hand, saving leisure until work is finished may facilitate working; people simply may find it easier to work if they keep a light at the end of the tunnel. A rich literature has examined the effects of imminent rewards on motivation and performance (see Cerasoli, Nicklin, & Ford, 2014). In the current studies, we explored another possibility: Saving leisure until work is finished may also reflect how people think about the *reward*. That is, people may assume that consuming "unearned" enjoyment—leisure first—would undermine its hedonic value and render the experience less pleasurable than if it took place after work was completed. Indeed, preferences for saving positive events and sensitivities to deservingness may lead people to assume that looming work would make an unwelcome distraction: People may assume that, rather than being immersed in the enjoyable leisure, they will be preoccupied with work tasks not yet completed (e.g., ruminating on how to execute the upcoming tasks well, dreading having to still do the tasks, concerned that current leisure is not sufficiently justified). Anticipated dread

**Corresponding Author:**
Ed O'Brien, Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL 60637
E-mail: eob@chicagobooth.edu

and related worries frequently shape how people think about future experiences (see Harris, 2012).

Such assumptions, however, may also often be mistaken. People have notoriously poor insight into how their attention will be divided by a multitude of stimuli (Kahneman & Thaler, 2006; Koehler, 1991), especially if that multitude includes hedonic states and experiences. For example, subjects in one study predicted that their enjoyment for snacking on chips would be disrupted after various distractor objects were placed nearby; however, the chips remained just as tasty in the presence of distractors (Morewedge, Gilbert, Myrseth, Kassam, & Wilson, 2010). Salty chips are, after all, salty chips; in the moment of munching, it can be hard to think of much else. Hsee and Zhang (2004) highlighted the related experience of comparison shopping. When comparing stereo speakers, for example, many shoppers think small shortcomings of inferior models will continue to bug them long after purchase; but once the speakers are humming along at home, any one model often provides just as much listening pleasure as the next.

More broadly, "hot" affective and hedonic states, such as the feelings of stimulation, excitement, or relaxation that come with leisure, are extremely absorbing and tend to dominate one's mind in the moment of consumption, over and above many other competing claims on attention (Ariely & Loewenstein, 2006; Bornstein, 1989; Buechel, Zhang, Morewedge, & Vosgerau, 2014; Csikszentmihalyi, 1990; Loewenstein, 1996). To the extent that people do not fully account for this power of immersion—after all, feelings are fleeting, and mental simulations rarely match the richness of the real thing (Campbell, O'Brien, Van Boven, Schwarz, & Ubel, 2014; Gilbert & Wilson, 2007; Robinson & Clore, 2002; Van Boven, Loewenstein, Dunning, & Nordgren, 2013)—they might generally think that the order of activities affects hedonic enjoyment more than it actually does. People who merely think about leisure before work may assume the reward of leisure would be spoiled and fail to appreciate the absorbing pleasures that they may still experience while actually engaged in leisure before work (despite external factors—such as order—that seemed more salient at a distance than up close).

We explored these possibilities in four studies. We assessed diverse activities ranging from eating snacks and getting massages to taking midterms and other strenuous tests. First, we examined a widely held intuition that consuming leisure before work undermines enjoyment (Study 1). We then conducted three experiments testing the validity of this intuition: Is leisure actually less enjoyable when it precedes, rather than follows, the completion of work (Studies 2 through 4)? We hypothesized that people may fail to appreciate the full pleasure of leisure before work as experienced during the moment of consumption. Last, we looked for mediation-based (Study 3) and moderation-based (Study 4) evidence for perceptions (or misperceptions) of attention and immersion in driving this effect. We report all measures, manipulations, and exclusions for all studies.

## Study 1: Intuitions About How Work Affects Leisure

First, we tested whether people indeed hold the intuition that leisure will be less enjoyable if consumed before (rather than after) work is completed.

### *Method*

Study 1 consisted of four smaller tests in which we varied the designs and populations to help generalize the hypothesized effect, testing for a similar pattern across different parameters. We sampled as many subjects as resources and availability allowed and then used these results for power analyses in subsequent studies.

***Study 1a.*** First, we embedded an item within an unrelated study on memory taken by 150 subjects in our downtown laboratory (mean age = 35.37 years, *SD* = 13.38; 33.33% female; 30.70% White, 49.30% Black, 6.00% Asian American or Asian, 14.00% mixed or other ethnicity). Subjects from this pool were recruited from the local city community to participate in university studies in exchange for pay. The study took about 15 min, and subjects were paid $3.00.

The embedded item was as follows:

> This next question pertains to your personal views on hedonic enjoyment—things you find fun and enjoyable. There are no 'right/good' or 'wrong/bad' answers here. Please report your honest agreement with the following statement: Hedonic enjoyment would be spoiled (less fun/enjoyable) for me if engaged in before completing an effortful task than if engaged in after completing an effortful task.

Subjects rated their agreement on a scale from 1 (*completely disagree*) to 10 (*completely agree*). We hypothesized that ratings would fall significantly above the midpoint, which would suggest endorsement of this belief.

***Study 1b.*** Second, we embedded an item within a battery of unrelated tasks taken by 155 students (43.9% female; age and ethnicity not recorded) enrolled in a course required for a master's degree in business administration (M.B.A.). The battery took about 15 min to complete, and students were invited to complete it (privately, anonymously, and voluntarily) before the start of the quarter.

The embedded item read as follows:

> Rate your agreement with the following statement. There are no 'right/good' or 'wrong/bad' answers, so please report how you honestly feel: In general, a positive experience becomes less enjoyable if consumed before (rather than after) getting a negative experience over with.

Subjects rated their agreement on a scale from 1 (*definitely disagree*) to 5 (*neither disagree nor agree*) to 9 (*definitely agree*). We hypothesized that ratings would fall significantly above the midpoint, which would suggest endorsement of this belief.

***Study 1c.*** Next, we conducted a between-subjects test on Amazon Mechanical Turk (MTurk) using texts and measures that were more thorough than those in the first two tests. We recruited 160 subjects (mean age = 31.11 years, *SD* = 9.55; 37.50% female; 75.00% White, 5.00% Black, 10.60% Asian American or Asian, 9.40% mixed ethnicity or other ethnicity) to complete the study in exchange for $0.25. Subjects read the following prompt (our manipulation is reported in brackets):

> This is about 'hedonic enjoyment,' things in life that you might do for leisure, fun, relaxation, or pleasure. Imagine that there is an opportunity for you to engage in a hedonic experience (e.g., the opportunity to go to a certain show, concert, restaurant, or destination; the opportunity to use a certain product or gadget; some other desirable experience). By chance, it happens to time up with an annoying task you have to complete at your job, such that you'd have to 'cash in' on this experience [before you'll begin/after you'll finish] working on this task. If you did engage in this experience at this point, how good do you think the experience would be for you? Note: There are no 'right' or 'wrong' answers.

Subjects then predicted how enjoyable, pleasurable, fun, positive, and beneficial the experience would be; these items were presented in randomized order, and each was rated on a scale from 1 (*not at all*) to 10 (*extremely*). We hypothesized that subjects would predict the experience to be significantly worse when they imagined it occurring before the work task rather than after.

***Study 1d.*** Finally, we conducted another between-subjects study and recruited 200 MTurk subjects (mean age = 35.50 years, *SD* = 11.14; 59.50% female; 76.50% White, 7.50% Black, 7.00% Asian American or Asian, 9.00% mixed ethnicity or other ethnicity) in exchange for $0.20. Subjects read the following prompt (our manipulation is reported in brackets):

> Imagine you work for a company that demands a lot of your energy and effort; it's a tough job. Your vacation days are randomly assigned each month. As you can see, imagine you get assigned the following days for this month, which fall at the [start/end] of the month.

Below that, subjects saw an image of an unmarked calendar with either the first 2 weekdays or the last 2 weekdays of the month checked off in green, each creating a 4-day weekend. Then, they predicted how enjoyable, relaxing, satisfying, and valuable this vacation would be and how much it would restore them for work ahead. These items were presented in random order, and each was rated on a scale from 1 (*not at all*) to 7 (*completely*). We tested whether subjects would intuitively predict that the vacation would be less enjoyable when merely framed as falling at the start, before the month's work is complete (even though it could just as well be seen as falling after having finished the previous month's work).

## Results

For Study 1a and Study 1b, we tested whether responses to the survey items were significantly different from the value at the midpoint of the scale. The community subjects significantly endorsed (midpoint = 5.5) the belief that hedonic experiences would become less enjoyable if consumed before (rather than after) an effortful task ($M$ = 6.37, $SD$ = 3.02), $t(149) = 3.54$, $p = .001$, $d = 0.58$, 95% confidence interval (CI) for the difference = [0.39, 1.36]. The M.B.A. students significantly endorsed (midpoint = 5.0) the belief that positive experiences would become less enjoyable if consumed before getting a negative experience over with ($M$ = 7.12, $SD$ = 1.77), $t(154) = 14.90$, $p < .001$, $d = 2.40$, 95% CI for the difference = [1.84, 2.40].

For Study 1c, we collapsed the items ($\alpha = .97$) and ran an independent-samples $t$ test. Subjects believed that otherwise similar leisure would be significantly less enjoyable if cashed in before ($M$ = 6.17, $SD$ = 2.08) rather than after ($M$ = 7.66, $SD$ = 1.80) completing a work task, $t(158) = -4.85$, $p < .001$, $d = 0.77$, 95% CI for the difference = [0.88, 2.10]. Likewise, for Study 1d ($\alpha = .94$), subjects believed an otherwise similar vacation would be significantly less enjoyable at the start of a month (before completing the month's work; $M$ = 4.22, $SD$ = 1.62) rather than at its end (after completing the month's work; $M$ = 4.71, $SD$ = 1.61), $t(198) = -2.18$, $p = .031$, $d = 0.30$, 95% CI for the difference = [0.05, 0.95].

### *Discussion*

Our hypothesis was supported for each of these tests. People appear to hold a general intuition that leisure would be spoiled if taken before they finished work as opposed to after they finished work.

One limitation of these initial tests is that agreement with survey items may partly reflect acquiescence. Moreover, we do not know the work and leisure experiences that subjects brought to mind. For example, some leisure is objectively spoiled if consumed first (e.g., filling up on celebratory beers just before running a 5-km race), whereas other leisure is not even possible until work is finished (e.g., waiting for an end-of-year bonus to fund a vacation). Some subjects could have imagined such examples, which may have led them to underappreciate leisure before work for perfectly good reasons beyond underappreciating the experience per se. Therefore, throughout Studies 2 through 4, we tested whether we could replicate these findings using clearly specified activities and tested whether intuitions about the order of these activities were indeed mistaken. Subjects were randomly assigned to serve as *predictors* (who imagined engaging in a leisurely activity either before or after completing a work task and estimated their reactions) or *experiencers* (who actually engaged in the activity in one of the two orders and reported their reactions). This allowed us to experimentally test the accuracy of people's intuitions about exactly the same activity, simply experienced at different times.

This experimental approach is not without trade-offs. The existence of the intuition implies that few study subjects would freely choose leisure first, which would lead to substantial attrition or selection concerns (e.g., the leisure-before-work subjects who do stay in such a study may have simply rescheduled their work to allow for the leisure, which defeats the purpose of the manipulation). Hence, in Studies 2 through 4, we assigned and enforced task order ourselves. In turn, the kinds of tasks that we can enforce as experimenters may be less personally consequential than some real-world tasks, given various practical constraints (e.g., lacking the managerial power to alter actual work schedules) and, not least, ethical constraints (e.g., being mindful about forcing people against their intuitions to rearrange tasks with real workplace consequences, especially before testing via controlled experiments whether scaling to such contexts would be worthwhile).

We used a variety of stimuli and methods to maximize external validity under these constraints. Future research should build on these studies with the goal of scaling to more naturalistic settings. Nonetheless, one general point to highlight up front is that although the constraints of the laboratory may limit generalizability from the experiencer's perspective (e.g., leaving highly personal work undone may spoil leisure to a greater degree than the work tasks in our studies), they cannot account for discrepancies in prediction. That is, predictors in Studies 2 through 4 know these constraints too and can presumably adjust their predictions accordingly, all else being equal. The critical test of our hypothesis is that, when people imagine specified tasks as consumed in specified contexts, they may fail to simulate their full enjoyment and immersion during leisure before work; they may generally believe that the surrounding context of a fun activity will matter more than it actually does once the activity is live and under way in real time.

## Study 2: The Magic-Maker and Fixed-Labor Tasks

In Study 2, subjects imagined or actually completed a fun task (what we called the Magic-Maker game) and a work task (what we called the Fixed-Labor task). We hypothesized that predictors might assume the game would be less enjoyable if played before (rather than after) getting work over with, whereas experiencers might enjoy the game similarly regardless of when they play it.

### *Method*

Sample size was predetermined using the effect sizes from Study 1 as estimates. For a more conservative test, we averaged the *d* values (*d*s = 0.58, 0.77, and 0.30) without including the extreme value (*d* = 2.40), which resulted in a *d* of 0.55. Because Study 2's critical test was whether there was a 2 × 2 between-subjects interaction, we converted this *d* value to the appropriate unit (*f* = .27; Cohen, 1988) and conducted a power analysis (G*Power 3: Faul, Erdfelder, Lang, & Buchner, 2007) of a four-group design using an *f* of .27, an $\alpha$ of .05, and a *df* of 1. This produced a recommended total sample size of 181 needed to achieve 95% power for detecting the critical interaction.

***Subjects.*** We recruited 181 visitors at the Museum of Science and Industry in Chicago, Illinois (mean age = 36.13 years, *SD* = 16.80; 58.60% female; 71.80% White, 6.60% Black, 7.20% Asian American or Asian, 14.30% mixed ethnicity or other ethnicity) to complete the study in private individual sessions in exchange for a small snack. We sent our own research assistants to the museum independently (with museum approval), which allowed us to collect the recommended sample size precisely.

***Procedure.*** Subjects were randomly assigned to condition in a 2 (role: experiencer, predictor) × 2 (order of leisure: before work, after work) between-subjects study

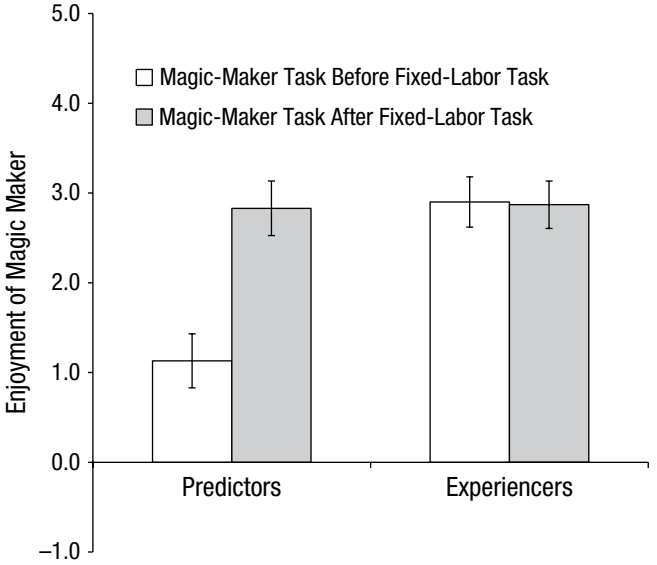


**Fig. 1.** Results from Study 2: mean enjoyment ratings for the Magic-Maker task (leisure) when it was played before and when it was played after the Fixed-Labor task (work), presented separately for predictors and experiencers. Error bars indicate ±1 *SE*.

about emotions. First, all subjects saw the activity materials and descriptions of the activities. The leisure task, Magic Maker, was described as follows:

> Many people report that the Magic Maker task is fun, interesting, and relaxing; generally speaking, it is a rather positive experience you will likely enjoy doing. In the task, you will play on the computer and seamlessly create beautiful pieces of music with simple touches on the screen. The task will last about 5 minutes.

This task was a level of the video game *Touch Pianist*. With each touch of various colors and shapes moving across the screen, players hear the notes of Beethoven's *Moonlight* Sonata play out one by one. No music skills are required. We loaded the game onto an iPad and attached a pair of headphones. The Fixed-Labor task, which we created, was described as follows:

> Many people report that the Fixed Labor task is dull, strenuous, and frustrating; generally speaking, it is a rather negative experience you will likely struggle through. In the task, you will complete a variety of math problems, word problems, and other calculations. The task will last about 5 minutes.

We printed a hard-copy packet of the task and attached it to a clipboard with a pen.

Some subjects then actually completed both tasks. To reduce suspicion, we told them the order would be based on an even draw from a deck of playing cards. Those who drew a black card had to complete the Magic-Maker task before the Fixed-Labor task and those who drew a red card had to complete the Magic-Maker task after the Fixed-Labor task. And this was true: They drew a card at random and completed the tasks in the preset order. After each task, the subjects rated their reactions by responding to the following measures on 11-point scales: "How much did you dislike-versus-enjoy this task?" (–5 = *extremely disliked*, +5 = *extremely enjoyed*), "How much displeasure-versus-pleasure did you feel during this task?" (–5 = *I felt displeasure the entire task*, +5 = *I felt pleasure the entire task*), and "Overall, how negative-versus-positive was this task for you?" (–5 = *completely negative*, +5 = *completely positive*).

Other subjects imagined going through all of these procedures, there and then, and predicted their reactions in one of the two orders. Predictors and experiencers were provided with exactly the same descriptions of the tasks.

## Results

For our primary analysis, ratings for the Magic-Maker task were collapsed into a scale ($\alpha$ = .85), and we conducted a univariate general linear model (GLM) with role, order, and the Role × Order interaction as independent variables and scores on this scale as the dependent variable.[1]

There was an incidental main effect of role such that predictors generally underestimated enjoyment, $F(1, 177) = 9.91$, $p = .002$, $f = .24$, and a main effect of order such that the game was indeed rated as less enjoyable when it came before, rather than after, the work task was completed, $F(1, 177) = 8.49$, $p = .004$, $f = .22$. This was qualified by the critical interaction, $F(1, 177) = 9.03$, $p = .003$, $f = .23$ (see Fig. 1). Pairwise comparisons revealed that predictors drove this effect: They thought the Magic-Maker task would be less enjoyable before ($M = 1.13$, $SD = 2.20$) rather than after ($M = 2.83$, $SD = 1.76$) getting the Fixed-Labor task over with, $F(1, 177) = 15.88$, $p < .001$, $d = 0.85$, 95% CI for the difference = [0.86, 2.55]. These results replicate those of Study 1. But as we hypothesized, this difference was less pronounced in reality: The game was similarly enjoyable whether played before ($M = 2.90$, $SD = 1.49$) or after the task ($M = 2.87$, $SD = 2.14$), $F(1, 177) = 0.01$, $p = .940$, $d = 0.02$, 95% CI for the difference = [–0.77, 0.73].

It is also informative to analyze the data within order (i.e., to compared predicted enjoyment with actual enjoyment when leisure came before work and compare predicted enjoyment with actual enjoyment when leisure came after work). Pairwise comparisons in this direction

revealed that predictors significantly underestimated their enjoyment when the Magic-Maker task was played before the Fixed Labor task, $F(1, 177) = 18.55$, $p < .001$, $d = 0.94$, 95% CI for the difference = [0.96, 2.58]; however, there was no significant difference between predicted and actual enjoyment when the Magic-Maker task was played after the Fixed-Labor task, $F(1, 177) = 0.01$, $p = .919$, $d = 0.02$, 95% CI for the difference = [–0.75, 0.84]. The misprediction indeed appears to have been driven by an assumption that unearned enjoyment would be spoiled and not that earned enjoyment would be boosted. People mistakenly believe that they would not be able to enjoy leisure as much as they normally would if work were left undone—at least when thinking about these tasks in these contexts.

These findings replicate and extend the basic effect in Study 1. The design of this study afforded strong control over the stimuli and procedures. We compared predicted ratings with actual ratings of the same well-defined tasks while ensuring that all subjects completed the tasks as intended. However, this also limited generalizability because the tasks were rather short, impersonal, and inconsequential. Thus, our next study was designed in hopes of replicating and extending this effect by exploiting the natural timing of a real-world work task.

## Study 3: Going to the Spa Before Rather Than After Midterms

In Study 3, students imagined or actually consumed a spa experience near the start or end of the campus midterm period. We hypothesized that experiencers might still enjoy the spa even with midterms looming, whereas predictors might assume it would be spoiled.

We also assessed mechanisms by having subjects report on various measures of attention during the spa experience. We hypothesized that mispredictions of enjoyment might be driven by corresponding mispredictions about distraction from looming midterms.

### *Method*

Sample size was predetermined using the effect size from Study 2 as an estimate, given that the current study follows the same statistical design. The same power analysis was conducted, except that $f$ was .23 (taken from Study 2). This produced a recommended total sample size of 248 needed to achieve 95% power for detecting the critical interaction.

To recruit these subjects, we launched the study through our subject pool around the campus midterm period, about a 3-week block in the middle of the quarter during which classes are expected to assign a midterm (typically an exam, a paper, or both). We built a "spa" in a private laboratory room, and subjects were offered a relaxing spa experience. Our goal was to naturalistically group subjects according to whether the majority of their own midterms were already finished (leisure after work) or still remaining (leisure before work) at the time they completed the study.

People in the subject pool at our university can show up to the laboratory by their own volition without signing up for times in advance, which made it difficult to predetermine how many people would show up for studies during the limited window of time during midterms. In addition, we did not advertise our study as being related to midterms, so people who did not meet the criteria (i.e., students who happened to not have any midterms or had just as many midterms done as remaining) could also sign up. We had to balance these concerns with our goal to reach the recommended sample size. Thus, we decided to first run only the experiencer conditions to maximize power. We intentionally oversampled within this window and requested that the laboratory run as many subjects as possible, and we monitored the number of premidterm sign-ups as a benchmark for postmidterm sign-ups. This resulted in 163 experiencers, of whom 131 met the criteria. We then launched the predictor conditions in the same population (but with unique individuals) on the basis of the experiencer sample. For simplicity in scheduling, we requested that the laboratory run at least 120 subjects rather than 131 precisely. We were informed when 128 predictors had been run (after the lab closed for a weekend), and we opted to stop there. This makes $N = 259$ for the current sample. The data file includes this sample plus the ineligible experiencers.

Earlier, we highlighted practical and ethical concerns about manipulating personal work tasks. The design of Study 3 circumvented these issues by exploiting the timing of naturally occurring work that students were likely to find important, rather than forcing one on them. However, the nonrandom assignment within the experiencer conditions raised concerns about selection effects (e.g., perhaps the subjects who showed up before finishing midterms were less distracted by midterms to begin with). There are similar tradeoffs regarding our decision to maximize power by running experiencers before predictors (e.g., perhaps the time of day that was available to experiencers and predictors was different, given that they were recruited at different parts of the academic calendar). This is an important concern. Study 3 did not use full random assignment, so we ultimately could not rule out potential selection effects that could affect spa enjoyment. However, various features lowered the possibility of major selection issues.

First, subjects at all stages were university students, recruited from the same population, in the same way, in

the same lab space, and within a short time span of the same quarter. Our subjects were paid, and they were not all enrolled in a particular course. The laboratory is open only during weekday afternoons (12:00 p.m. to 6:00 p.m.). There were no systematic differences in the composition of any of the four cells in terms of subjects' sex, age, and ethnicity or the day of the week and time of day that they completed the study (see the Supplemental Material available online).

Second, there is no evidence that subject-pool traffic dropped around midterms. All available traffic data (provided by the lab staff) revealed a steady flow throughout the quarter in which our study was conducted, and this held true for each midterm period over the preceding year of archived data (see the Supplemental Material). This reduced concerns about substantive shifts in the kinds of subjects who show up for studies over time. Again, the pool was public and pay based, so students may have participated for many reasons beyond simply having free time in a course.

Third, recruiting all experiencers within the midterm period helped reduce the likelihood that their construal of midterms differed in incidental ways at the time of the spa. For example, all experiencers presumably had midterms on their minds in similar ways (e.g., being surrounded by the same campus reminders, not yet knowing all their grades), and the critical difference was a different remaining workload. Note also that our premidterm experiencers did have many midterms left to go, which suggests some degree of distraction (presumably the freest and calmest students were the students without any midterms, but they were excluded). We directly address this concern in the Results section.

Finally, predictors also knew the features of the study: They were asked to imagine freely coming to the laboratory during the midterm period having either completed or not yet completed the majority of their midterms. This allowed predictors to adjust their predictions accordingly. As outlined earlier, the critical test for Studies 2 through 4 is whether predictions of these specified study contexts are miscalibrated.

***Subjects.*** We recruited 259 subjects from our campus subject pool (mean age = 19.45 years, *SD* = 1.77; 45.60% female; 48.60% White, 9.70% Black, 22.80% Asian American or Asian, 18.90% mixed ethnicity or other ethnicity) to complete the study in private individual sessions in exchange for $3.00.

***Procedure.*** Subjects were quasirandomly assigned to condition in a 2 (role: experiencer, predictor) × 2 (order of leisure: before work, after work) between-subjects design. As outlined, the order variable for experiencers was defined by whether they participated in the study before or after completing the bulk of their midterms. Students in the before-work group, on average, had already taken 1.77 midterms (*SD* = 1.61) and had 4.17 left (*SD* = 1.96), whereas those in the after-work group, on average, had already completed 4.06 midterms (*SD* = 1.72) and had 1.33 left (*SD* = 1.53).

Experiencers were brought into the room one at a time. The room contained an electronic massage chair, an electronic footbath with fresh water and towel, relaxing artwork, and candles. The experimenter explained how to work these items and informed subjects that they would sit by themselves with calming music turned on and the lights turned off. Their task was to find the experimenter once the music stopped, after about 6 min. The experimenter then left the room and took the subject's belongings (including cell phones). After the subjects found the experimenter, they privately rated their experience of the spa on a computer by responding to the same enjoyment items used in Study 2.

After the experiencers rated their enjoyment, the survey continued to a new screen that asked about their attention and distraction during the spa experience: "Across the duration of the spa experience itself, what percentage of your 'mind' in total would you say reflected each of the following categories?" They saw four categories and had to type a percentage for each; the percentages for the four categories had to sum to 100%. The four categories were "% spent distracted by negative things in my life these days, like midterms and school stress" (critical category of interest); "% spent tuned out and just experiencing the spa experience itself"; "% spent distracted by positive things in my life these days, other than the spa experience"; and "% spent distracted by other things that don't fit into the categories here."

We expected to find differences on "negative things like midterms" such that predictors would overestimate the extent to which midterms actually dominate attention during the spa experience. It was less clear how these ratings should disperse into the remaining categories: Because the items were interdependent, subjects who gave more weight to midterms must have given less weight somewhere else, but each subject might allocate differently. The critical test of our hypothesis depended on the absolute percentage of attention absorbed by midterms.

Finally, experiencers were asked to report the number of midterms that they had "already taken or already completed so far this quarter" and also the number of midterms that they "still have left to take or complete this quarter."

As in Study 2, other subjects served as predictors. Predictors imagined that they "came into the lab to participate in studies during the point of the quarter around midterm exams" and imagined coming in either before or

after completed the majority of their midterms. They were given the same exact descriptions of the tasks that the experiencers were given, and they were asked to predict their enjoyment and attention using the same scales that the experiencers used.

## *Results*

***Mispredicting enjoyment.*** For our primary analysis, the enjoyment ratings for the spa experience were collapsed into a scale ($\alpha = .93$), and we conducted a univariate GLM with role, order, and the Role × Order interaction as independent variables and scores on this scale as the dependent variable.

There was an incidental main effect of role such that predictors generally underestimated enjoyment, $F(1, 255) = 16.98$, $p < .001$, $f = .26$, and a main effect of order such that the spa was indeed rated as less enjoyable before rather than after completing the bulk of their midterms, $F(1, 255) = 30.33$, $p < .001$, $f = .34$. This was qualified by the critical interaction, $F(1, 255) = 32.93$, $p < .001$, $f = .36$ (see Fig. 2). Pairwise comparisons revealed that predictors drove this effect: They thought the spa would be less enjoyable before ($M = 1.47$, $SD = 2.27$) rather than after ($M = 3.84$, $SD = 1.32$) completing the bulk of their midterms, $F(1, 255) = 62.50$, $p < .001$, $d = 1.27$, 95% CI for the difference = [1.78, 2.96], which replicates the results of Studies 1 and 2. As with the experiencers in Study 2, however, this difference was less pronounced in reality: The spa was similarly enjoyable before ($M = 3.55$, $SD = 1.13$) and after ($M = 3.50$, $SD = 1.84$) completing the bulk of their midterms, $F(1, 255) = 0.03$, $p = .870$, $d = 0.03$, 95% CI for the difference = [–0.63, 0.54]. A massage is a massage; in the moment, when people get it may matter less.[2]

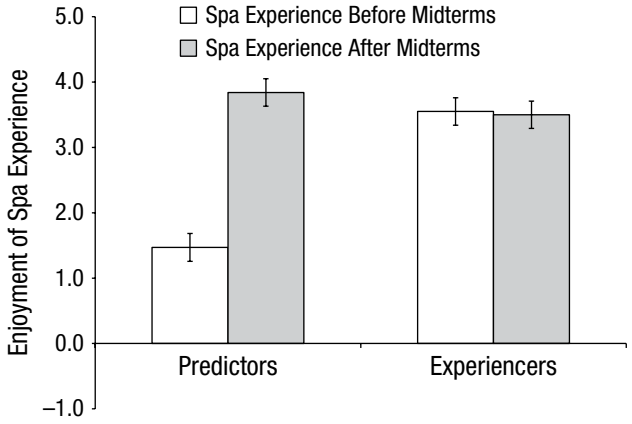


**Fig. 2.** Results from Study 3: mean enjoyment ratings for the spa experience (the leisure task) when it came before and when it came after students had completed most of their midterms (the work task), presented separately for predictors and experiencers. Error bars indicate ±1 *SE*.

Likewise, when we analyzed the data within order (i.e., comparing predicted enjoyment with actual enjoyment when leisure came before work and comparing predicted enjoyment with actual enjoyment when leisure came after work), pairwise comparisons in this direction again demonstrated that predictors significantly underestimated their enjoyment when the spa was experienced first (before completing the bulk of their midterms), $F(1, 255) = 48.04$, $p < .001$, $d = 1.16$, 95% CI for the difference = [1.49, 2.67]; however, when the spa was experienced second (after completing the bulk of their midterms) predicted and actual enjoyment did not significantly differ, $F(1, 255) = 1.33$, $p = .251$, $d = 0.21$, 95% CI for the difference = [–0.93, 0.24]. As in Study 2, people appear to believe, mistakenly, that leisure before work would be spoiled, rather than thinking that leisure after work would be boosted.

***Mispredicting attention toward midterms.*** Subjects' misprediction of enjoyment may reflect the corresponding misprediction that thoughts about their midterms would distract them during the spa experience. To test this possibility, we conducted a multivariate GLM with role, order, and the Role × Order interaction as independent variables and the four percentage-estimate categories—attention toward "negative things like midterms" (critical variable of interest), "the spa experience itself," "unrelated positive things," and "other thoughts"—as the dependent variables.

Descriptive statistics for each category are reported in Table 1. First and foremost, we found all hypothesized effects for "negative things like midterms." There was a main effect of role, $F(1, 255) = 16.79$, $p < .001$, $f = .26$; predictors generally overestimated the extent to which they would be paying attention to their midterms. There was also a main effect of order, $F(1, 255) = 43.91$, $p < .001$, $f = .41$; subjects indeed reported being more distracted by midterms before, rather than after making it through the bulk of them. These main effects were qualified by the critical interaction, $F(1, 255) = 8.68$, $p = .004$, $f = .19$: Although experiencers were indeed more distracted by midterms before compared with after completing the bulk of them, $F(1, 255) = 6.86$, $p = .009$, $d = 0.44$, 95% CI for the difference = [2.05, 14.52], predictors significantly overestimated this difference, $F(1, 255) = 45.28$, $p < .001$, $d = 1.23$, 95% CI for the difference = [15.24, 27.85]. This also acted as a manipulation check of sorts. Midterms were more distracting during the spa experience when experiencers had many midterms left to complete than when they had already finished most of their midterms;[3] as we hypothesized, however, the experiencers' actual distraction was not as great as the predictors

**Table 1.** Results From Study 3: Mean Percentages Assigned to Each Category of Attention

| | Predicted attention and immersion during the spa experience | | Actual attention and immersion during the spa experience | |
|---|---|---|---|---|
| Category | Spa before bulk of midterms | Spa after bulk of midterms | Spa before bulk of midterms | Spa after bulk of midterms |
| Negative things like midterms* | 35.73% (21.75%) | 14.18% (11.93%) | 19.88% (22.08%) | 11.59% (14.68%) |
| Spa experience itself | 35.52% (21.75%) | 48.66% (23.89%) | 43.89% (25.95%) | 47.35% (25.41%) |
| Unrelated positive things | 11.65% (9.53%) | 19.74% (14.06%) | 16.92% (14.15%) | 19.77% (15.85%) |
| Other thoughts | 17.10% (16.29%) | 17.42% (17.93%) | 19.31% (19.08%) | 21.29% (21.07%) |

Note: Values in parentheses are standard deviations. Subjects were asked the following question: "Across the duration of the spa experience itself, what percentage of your 'mind' in total would you say reflected each of the following categories?" They then assigned percentages to the four categories, and the percentages had to sum to 100%. The asterisk indicates the only significant Role × Order interaction ($p = .004$).

assumed and was not enough to disrupt enjoyment of the spa experience.

As in Study 2, pairwise comparisons within order again showed that subjects thought that their enjoyment of the spa experience would be spoiled if it occurred before they completed the bulk of their midterms rather than that their enjoyment of the experience would be boosted if it occurred after they completed the bulk of their midterms: Subjects significantly overestimated the degree to which attention to their midterms would distract them from experiencing the spa before completing the bulk of their midterms, $F(1, 255) = 24.51$, $p < .001$, $d = 0.72$, 95% CI for the difference = [9.55, 22.16]; however, when students experienced the spa after completing the bulk of their midterms, predicted and actual distraction did not differ significantly, $F(1, 255) = 0.67$, $p = .413$, $d = 0.19$, 95% CI for the difference = [–3.64, 8.83].

Regarding the remaining three categories, for ratings of the "spa experience itself," there was no main effect of role, $F(1, 255) = 1.36$, $p = .244$, $f = .07$, but there was a main effect of order, $F(1, 255) = 7.53$, $p = .007$, $f = .17$; subjects indeed reported a harder time immersing themselves into the spa experience before, rather than after making it through the bulk of their midterms. The interaction did not reach statistical significance, $F(1, 255) = 2.56$, $p = .111$, $f = .10$. Likewise, for ratings of "unrelated positive things," there was no main effect of role, $F(1, 255) = 2.45$, $p = .119$, $f = .10$, but there was a main effect of order, $F(1, 255) = 10.41$, $p = .001$, $f = .20$; subjects reported being less mindful of unrelated positive things when the spa experience occurred before, rather than after, they completed the bulk of their midterms. The interaction again did not reach statistical significance, $F(1, 255) = 2.39$, $p = .124$. Finally, there were no differences at all for ratings of "other thoughts": There was no main effect of role, $F(1, 255) = 1.71$, $p = .192$, $f = .08$; no main effect of order, $F(1, 255) = 0.25$, $p = .621$, $f = .03$; and no interaction, $F(1, 255) = 0.13$, $p = .721$, $f = .01$.

The pairwise analyses for each of these categories are provided in the Supplemental Material. The analyses for "spa experience itself" and "unrelated positive things" may be most relevant: Predictors believed that they would struggle to focus on these positive categories before rather than after completing the bulk of their midterms, $F(1, 255)$s $\geq 9.33$, $p$s $\leq .002$, $d$s $\geq 0.58$, whereas actual attention did not significantly differ before and after, $F(1, 255)$s $\leq 1.43, 9.33$, $p$s $\geq .233$, $d$s $\leq 0.19$. This is conceptually consistent with our hypothesis, but we hesitate to draw further inferences because the interactions did not reach statistical significance.

Again, the significant interaction and pairwise effects for ratings of "negative things like midterms" is critically important and suggests that predictors overestimate the extent to which midterms dominate attention (regardless of how attention to other things may be divided).

***Mediation.*** To sum up our results so far, subjects who imagined having the bulk of their midterms left undone significantly overestimated the extent to which they would be distracted by midterms actually experiencing the spa.

To test whether this misprediction of attention accounted for why subjects mispredicted their enjoyment, we used the Process macro (Version 2.16; Hayes, 2013) for IBM SPSS to conduct mediation analyses according to the recommended procedures. We conducted a moderated mediation analysis (Model 58, 5,000 iterations) with order as the condition variable, role as the moderator, enjoyment scale as the dependent variable, and subjects ratings of attention to negative things as the mediator. Model 58 tests whether the effect of order on enjoyment was indeed mediated by attention to midterms (conditional indirect effects of attention on enjoyment) and how this process was affected by role, both on the path from order to attention to midterms (first-stage moderation) and on the path from attention to midterms to enjoyment (second-stage moderation).

First, the indirect effect of order on enjoyment, via attention to midterms, was significant for both predictors, indirect effect = 1.33, bootstrapped *SE* = 0.29, 95% bootstrapped CI = [0.80, 1.94], and experiencers, indirect effect = 0.11, bootstrapped *SE* = 0.07, 95% bootstrapped CI = [0.01, 0.29]. These findings suggest mediation: As hypothesized, subjects generally drew on midterm distraction to inform their spa enjoyment. Second, there was also evidence of significant moderated mediation, index = −1.22, bootstrapped *SE* = 0.30, 95% bootstrapped CI = [−1.84, −0.68]. This suggests that the difference between predictors and experiencers was significant: Predictors were more influenced by midterms than experiencers were. Finally, we also observed significant interactions at each stage: Role affected the path between order and midterms, $b = 13.26$, $SE = 4.50$, $t(255) = 2.95$, $p = .004$, and the path between midterms and enjoyment, $b = 0.05$, $SE = 0.01$, $t(254) = 4.71$, $p < .001$. This suggests that predictors were more influenced than experiencers in two ways, each of which amplified the mediating influence of midterms on enjoyment: Predictors overestimated the extent to which experiencers were distracted by midterms to begin with, as well as how damaging thoughts about midterms proved to be for spa enjoyment (for the complete output from the Process macro for our mediation analysis, see the Supplemental Material).

Study 3 replicated and extended the basic effect using more realistic work and leisure stimuli and also provided evidence for an attentional mechanism. People did not fully appreciate the enjoyment they could reap from leisure before work, perhaps because they overestimated the extent to which the upcoming work would be distracting.

Our moderated mediation results support such a process. First, we confirmed a direct link between being distracted by midterms left undone and struggling to enjoy the spa, which suggests mediation by attention as hypothesized. Second, predictors overestimated the strength of this link: They overestimated not only the presence of distractions but also the extent to which such distractions spoiled people's ability to enjoy themselves anyway. Both possibilities suggest that people generally underappreciate the full extent of their immersion into a fun activity during the moment of consumption.

In our next study, we sought to extend these findings further. If a lack of appreciation for attention and immersion drives the effect, then reminding people about the power of immersion should reignite (and therefore calibrate) their dulled intuitions about leisure before work.

## Study 4: Appreciating Hedonic Immersion

In Study 4, subjects had to consume their reward for completing the strenuous "Cognitive Marathon" before they actually completed it. We hypothesized that the reward would still be enjoyable, contrary to predictions. However, predictions might improve after a debiasing intervention on the power of immersion.

### Method

Sample size was predetermined using the effect sizes from Study 1 as estimates, given that the current study tests for differences among three conditions rather than a 2 × 2 interaction, as in Studies 2 and 3. A power analysis was conducted using a $d$ of 0.55 (the average effect size from Study 1, again excluding the extreme value), two-tailed, $\alpha$ of .05, and a ratio of 1. This produced a recommended total sample size of 261 to achieve 95% power for detecting group differences. For simplicity in scheduling and to maximize power, we requested that the laboratory run about 300 subjects. This was reached in the middle of a week, and the study was kept live for the week's remainder to fulfill extant sign-ups.

***Subjects.*** We recruited 332 subjects from our campus subject pool (mean age = 28.75 years, *SD* = 12.23; 46.40% female; 30.70% White, 39.80% Black, 14.50% Asian American or Asian, 15.00% mixed or other ethnicity) to complete the study in private individual sessions in exchange for $3.00. This study was open to all members of the campus community (not just undergraduates).

***Procedure.*** Subjects were randomly assigned to one of three conditions (experiencer, control predictor, or predictors who unpacked the scale of their attention) in a between-subjects study.

First, all experiencers were told that they would complete a study on work and cognitive performance. We called this the *cognitive marathon*, a strenuous battery of "puzzles, logic games, math problems, and other tasks, which range from computer-based survey questions, to pen-and-paper calculations, to playing around with physical objects to find a solution (e.g., similar to a Rubik's Cube)." They were told that everyone could and would successfully finish but it was a marathon in that they would likely feel drained when they were done. Because of this, we told experiencers that we had teamed up with other researchers who happened to be running an enjoyment study at the same time, so they would then take this study as a reward for completing our cognitive marathon.

Next, as experiencers waited in the study room, the experimenter returned and announced that some of the materials for the cognitive marathon were accidentally still in use and that, in order not to interfere with the laboratory's timing and scheduling, they would need to complete the enjoyment study first. Experiencers then completed the enjoyment study: They picked a snack from a bowl of choices (chocolates, pretzels, and popcorn) and, while eating, watched a 3-min compilation of funny

YouTube videos used in research on enjoyment (O'Brien, 2015). After watching, they rated their "video and snacking experience" privately on a computer via the enjoyment items used in Studies 2 and 3. On the next screen, they rated their immersion via a single global item, "During the task itself, to what extent were you 'immersed' in the video + snacking experience, i.e., able to just experience and enjoy it?" using a scale from 1 (*not at all immersed*) to 11 (*totally immersed*). Finally, we probed suspicion by asking experiencers the following question:

> At this point of the study, you probably figured out that there is no Cognitive Marathon (rest assured you will still be paid the full $3.00!). However, we'd like to ask: As you were going through the study, to what extent did you believe that you WOULD complete the Cognitive Marathon later on?

They had to choose between two options: "Yes, I believed—As I was going through the study today, I did believe that I would complete the Cognitive Marathon later on" and "No, I didn't believe—As I was going through the study today, I knew the Cognitive Marathon was fake and that I wouldn't actually complete it later on." After choosing, experiencers were fully debriefed (via a written form) regarding the deception and purpose of the study.

This deception served as another way to approach our hypothesis while operating under the outlined constraints. We were able to simply frame the work as weighty and substantial (without having to then impose this level of labor in reality or impinge on personally weighty tasks) and the leisure as a direct reward for completing it, which may have put subjects in a mind-set that may better capture what their natural experience would be like when consuming leisure while the hard work to which it is tied remains looming.

As in our previous studies, other subjects served as predictors. Predictors in this study were shown all of the original prompts and questions that experiencers saw. They were informed that other subjects served as experiencers who ended up having to complete the enjoyment study before, rather than after, completing the cognitive marathon. They were asked to estimate both the enjoyment and immersion of the average experiencer using the same items to which the experiencers responded.

Other subjects also served as predictors, but they first completed a debiasing intervention. We based our intervention on previous research on defocusing, in which inducing people to break down, or unpack, their attention into smaller units reduced various forecasting errors (Ayton, Pott, & Elwakili, 2007; Hoerger, Quirk, Lucas, & Carr, 2010; Van Boven & Epley, 2003; Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000). For example, Wilson et al. (2000) found that students overestimated how negative they would feel in the days after their university's football team suffered a loss; however, asking students to first list exactly what they might be doing at every hour for each of these 24-hr periods reminded them of all the things beyond football that would inevitably steal their attention, which led them to calibrate their predictions accordingly.

We designed a similar unpacking task at the scale of second-by-second attention to match the current study. These predictors first read about the experiencer procedures in full and completed all prompts as control predictors. But before making their predictions, they responded to a set of general questions about laughing-related and eating-related experiences and sensations. First, they unpacked laughing; these items consisted of a forced-choice question, a listing task, and a percentage estimate.

The forced-choice question was "Think about the experience of laughter, while in the moment of actually watching a funny video. What best describes this in-the-moment experience?" Subjects chose between two responses: "Mostly just watching and laughing—It just 'happens,' like a feeling or reaction" and "Mostly active work—You have to actively make it 'happen.'" We presumed that most subjects would choose the former option.

Next, the listing task had the following prompt:

> Let's 'zoom in' on this moment, second by second. Imagine someone is watching a funny clip and they begin experiencing laughter and all of its feelings, pulses, breaths, and so on, which lasts for about 10 s. What is going on moment-to-moment during this experience?

Subjects saw nine lines, starting with "Sec 1 to Sec 2" and ending with "Sec 9 to Sec 10." For each, they chose between two options: "At this exact second, you're simply 'absorbed' in the experience regardless of other thoughts" and "At this exact second, you're actively fighting off other thoughts to make these sensations happen." We presumed that subjects would choose the former option for most of the lines.

Last, the percentage-estimate item read as follows: "When a person is right in the actual moment of watching a funny video, what 'percentage' of the present moment is typically absorbed by it right then and there?" Subjects responded using a slider on the screen, with anchors of 0% and 100% and 1% increments. We presumed that subjects, on average, would choose high estimates.

Subjects then moved on and completed the same set of unpacking questions, except that the questions were

adapted for the experience of taste. The corresponding prompts were as follows:

- "In general, think about the experience of taste, while in the moment of actually chewing on a tasty snack. What best describes this in-the-moment experience?" The subjects chose between "Mostly just chewing and tasting—It just 'happens,' like a feeling or reaction" and "Mostly active work—You have to actively make it 'happen'."
- "Let's 'zoom in' on this moment, second by second. Imagine that someone is eating a tasty snack and they begin experiencing all of its flavors, textures, tastes, and so on, which lasts for about 10 s. What is going on moment-to-moment during this experience?" The subjects responded using the same format as for the laughter question (i.e., second-by-second responses).
- "When a person is right within the actual moment of tasting a snack, what 'percentage' of the present moment is typically absorbed by it right then and there?" The subjects responded using the same format as for the laughter question (i.e., 0% to 100% slider).

After this unpacking task, which should have brought to mind the absorbing nature of hedonic experiences, these predictors were asked to estimate the enjoyment and immersion of the average experiencer in the same way that the control predictors did.

## Results

***Manipulation checks.*** The cover worked as intended. The vast majority of experiencers (91.90%, 102 of 111) believed the cognitive marathon was truly going to occur later on (logistic regression: $b = -2.43$, $SE = 0.348$, $p < .001$).

The debiasing intervention also worked as intended. For the laughter question, almost all of these predictors (92.90%, 104 of 112) believed the moment of laughter involved "Mostly just watching and laughing" (logistic regression: $b = -2.57$, $SE = 0.367$, $p < .001$). In the listing task, they reported a mean of 5.90 ($SD = 1.94$) "absorbed" choices, much higher than chance (i.e., 4.50), one-sample $t(111) = 7.66$, $p < .001$, $d = 1.45$, 95% CI for the difference = [1.04, 1.76]. For the percentage estimates, they reported a mean of 74.59% ($SD = 17.71\%$) of the moment of laughter is absorbed by the hedonic experience itself, much higher than chance (i.e., 50.00%), one-sample $t(111) = 14.69$, $p < .001$, $d = 2.78$, 95% CI for the difference = [21.27, 27.91]. For the taste question, the majority of these predictors (77.70%, 87 of 112) believed the moment of taste involved "Mostly just chewing and tasting" (logistic regression: $b = -1.25$, $SE = 0.227$, $p < .001$). In the listing task, they reported a mean of 6.20 ($SD = 2.29$) "absorbed" choices, which was much higher than chance (i.e., 4.50), one-sample $t(111) = 7.83$, $p < .001$, $d = 1.48$, 95% CI for the difference = [1.26, 2.13]. For the percentage estimates, they reported that a mean of 71.37% ($SD = 21.70\%$) of the moment of laughter would be absorbed by the hedonic experience itself, which was much higher than chance (i.e., 50.00%), one-sample $t(111) = 10.42$, $p < .001$, $d = 1.97$, 95% CI for the difference = [17.30, 25.43].

***Predictions versus reality (enjoyment).*** For our primary analysis, enjoyment ratings were collapsed into a scale ($\alpha = .87$), and we conducted an analysis of variance (ANOVA) with condition as the independent variable and this scale as the dependent variable.

There was the hypothesized omnibus effect of condition, $F(2, 329) = 3.27$, $p = .039$, $f = .16$ (see Fig. 3). First, planned contrasts revealed that control predictors ($M = 2.52$, $SD = 1.73$) significantly underestimated experienced enjoyment ($M = 3.06$, $SD = 1.81$), $t(329) = 2.42$, $p = .016$, $d = 0.31$, 95% CI for the difference = [0.07, 1.01]. This finding replicated the basic effect observed throughout our previous studies. But more critical for the current study was that predictors who unpacked the scale of their attention were indeed more accurate: Unpacked predictions ($M = 2.95$, $SD = 1.45$) did not significantly differ from actual enjoyment, $t(329) = .48$, $p = .630$, $d = 0.07$, 95% CI for the difference = [–0.32, 0.54]. In turn, control predictors significantly underestimated enjoyment compared with unpacked predictors, $t(329) = -1.94$, $p = .043$, $d = 0.27$, 95% CI for the difference = [0.01, 0.86]. Bringing to mind the power of immersion reduced the pervasive prediction error.

***Predictions versus reality (immersion).*** We conducted an ANOVA with condition as the independent variable and the immersion item as the dependent variable. All results replicated the patterns that we observed for enjoyment.

We found the hypothesized omnibus effect of condition, $F(2, 329) = 4.35$, $p = .014$, $f = .19$. Planned contrasts revealed that control predictors ($M = 7.01$, $SD = 2.40$) significantly underestimated actual experienced immersion ($M = 7.89$, $SD = 2.46$), $t(329) = 2.85$, $p = .005$, $d = 0.36$, 95% CI for the difference = [0.24, 1.53], whereas predictors who unpacked the scale of their attention more accurately predicted experienced immersion ($M = 7.65$, $SD = 2.00$), $t(329) = .78$, $p = .435$, $d = 0.11$, 95% CI for the difference = [–0.35, 0.83]. In turn, control predictors significantly underestimated immersion compared with predictors who unpacked the scale of their attention, $t(329) = -2.16$, $p = .032$, $d = 0.29$, 95% CI for the

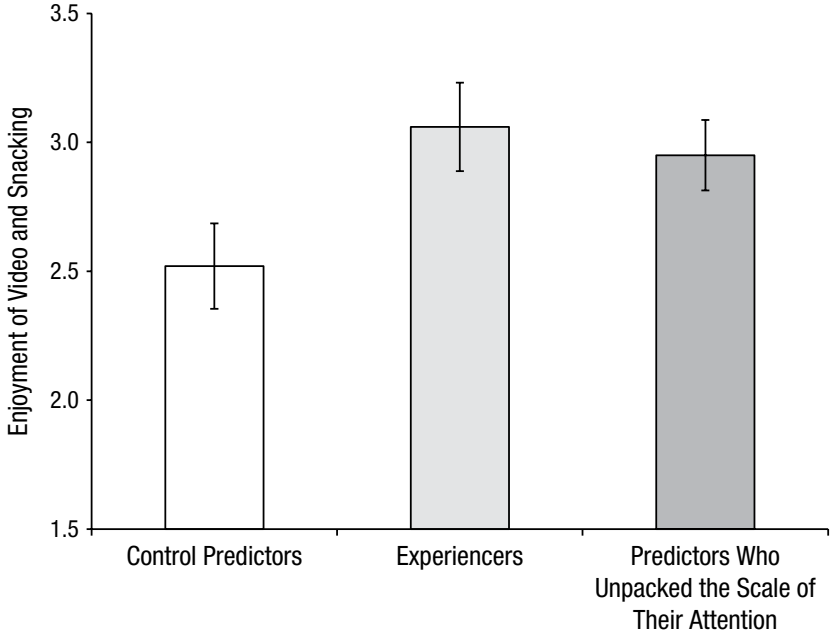


**Fig. 3.** Results from Study 4: mean enjoyment ratings for the video and snacking experience (the leisure task) when it took place before the cognitive marathon (work task), presented separately for control predictors, experiencers, and predictors who unpacked the scale of their attention. Error bars indicate ±1 *SE*.

difference = [0.06, 1.23]. Predictors underestimated their immersion by default, but came to better appreciate their immersion after unpacking the scale of their attention.

***Mediation.*** We tested for whether mispredictions of immersion drove mispredictions of enjoyment, which served as a replication of the mediation results of Study 3. We dummy-coded condition according to the guidelines for categorical variables that have three levels (Hayes, 2013): 0 = experiencer, 0 = predictor who unpacked the scale of his or her attention, and 1 = control predictor. We then conducted the recommended analyses (Model 4 at 5,000 iterations). The indirect effect of condition on enjoyment, via immersion, was significant, indirect effect = −0.35, bootstrapped *SE* = 0.13, 95% bootstrapped CI = [–0.62, –0.12], and the direct effect of condition on enjoyment was statistically reduced when controlling for immersion, $b = -0.14$, $SE = 0.15$, $t(329) = -0.89$, $p = .372$. This finding suggests that immersion indeed mediated the effect of the debiasing task on enjoyment. All results held when we reran these analyses between only the two predictor conditions, further suggesting that the debiasing intervention worked primarily by affecting beliefs about immersion (for the complete output from the Process macro for our mediation analysis, see the Supplemental Material).

These results replicated and extended our previous findings. People generally underestimate their enjoyment of leisure that takes place before work and assuming that work left undone would distract them during the leisure experience . Accordingly, inducing people to unpack the full power of immersion attenuates this effect and helps calibrate predictions.

## General Discussion

People generally believe that otherwise identical leisure would be undermined depending on when they consume it, that is, before (rather than after) getting work over with. Across various experiences, this intuition proved mistaken: Playing a fun game, receiving a relaxing massage, and indulging in snacks and fun videos were just as enjoyable and absorbing with work left undone, despite predictions that leisure would be spoiled.

These findings run counter to the presumed payoffs underlying the organization of work and leisure in everyday life. Many incentive systems depend on people believing that enjoyable rewards are best saved until work is finished; the implication that a reward is just as good before work as it would be after work could jeopardize people's willingness to do work in the first place (Cerasoli et al., 2014). Prescriptive insights therefore depend on one's goal. People may be able to promote delayed gratification by exploiting their intuition (like the framing in Study 1d); paradoxically, to improve employee attitudes about a distant reward, a manager could assign

a hard task sooner (prereward) rather than later (postreward), which makes the payoff appear bigger. But this strategy may hurt more than help if one's goal is to maximize real-time pleasure. Not only might people forgo leisure in the present, but waiting too long likely builds pressure to savor it, which is disruptive (Gruber, Mauss, & Tamir, 2011). It may sometimes be wisest to take leisure now if one's goal is to enjoy it. Serving this goal may even boost working. To the extent that people leave a movie feeling happy, a spa feeling relaxed, or an office party feeling bonded (without being spoiled by looming work, as our studies suggest), positive emotions at the outset make some work tasks easier to endure (Lyubomirsky, King, & Diener, 2005). Indeed, breaks and recovery from work benefit workers in many ways, from increasing job satisfaction to increasing productivity (Sonnentag, 2012; Sonnentag & Fritz, 2015). If people intuitively avoid leisure first, they may take advantage of such opportunities too rarely, to their own detriment. Judgment aids like the unpacking task in Study 4 may help people fight these intuitions.

These findings raise many interesting questions. To what extent and in what ways does the basic effect generalize? Future research should seek to scale the current studies and test whether leisure before work is still enjoyable in more naturalistic settings. One potentially important difference is that our studies may have minimized the role of guilt. Upcoming work tasks that are more personally consequential than our study tasks (e.g., the looming date of one's GRE exam) may seem more urgent to address and may therefore be especially distracting during preceding leisure opportunities (e.g., feeling obligated to mentally rehearse vocabulary words during a holiday, construing any time outside of studying as procrastination, or second-guessing one's choice to relax). Guilt is powerful (Kivetz & Simonson, 2002). Although upcoming work tasks can be distracting in many ways beyond guilt (Harris, 2012)—consider the mere dread or disappointment of seeing some annoying event pop up on the calendar—future research should disentangle whether different sources of distraction wield different influences on immersion and enjoyment.

Nonetheless, predictors in our studies imagined our specific study stimuli and still misperceived what these experiences were like, which suggests that people overgeneralize the intuition even when it is a poor approximation of the reality at hand (the hallmark of a bias; Baron, 1990). This discrepancy highlights two ways people may overgeneralize. First, people may overestimate the frequency of work-related distractions during leisure and assume they will be stuck thinking about work in situations in which they will not be thinking about work at all; people may imagine every day of vacation being spoiled by looming work, whereas just a handful may be spoiled in reality. Second, people may overestimate the influence of work-related distractions during leisure and assume that thinking about work will bother them to a greater degree than it does; the thought of the office back home may sting less while relaxing on a distant beach. Both possibilities run throughout our studies, particularly the mediation data in Study 3. Even if looming work distracts people more or differently than our stimuli did (e.g., because of strong guilt), we suspect that people may generally overestimate the frequency and influence of this distraction while having fun in the meantime.

Our consistent evidence that people believe that unearned enjoyment will be spoiled and do not believe that earned enjoyment will be boosted also raises broader implications. Although the completion of more personally consequential tasks may elicit a greater sense of accomplishment than our stimuli (which may indeed actively boost leisure after work), the same clear pattern of belief in "spoiling" across studies is noteworthy. These findings join several others (e.g., Buechel et al., 2014; Hsee & Zhang, 2004; Morewedge et al., 2010) that highlight the unforeseen power of hedonic experiences in absorbing attention within the moment of consumption. And yet, still other studies seem to highlight the power of the latter: Wine does taste cheaper when people think it is cheap (Plassmann, O'Doherty, Shiv, & Rangel, 2008), beer does taste nastier when people think it contains a nasty ingredient (Lee, Frederick, & Ariely, 2006), artwork does fail to stimulate when people think it is forged (Newman & Bloom, 2011), and so on. What explains the difference? Again, our critical effect is the misprediction, which reflects people's oversimplified simulations of rich hedonic states (Gilbert & Wilson, 2007; Robinson & Clore, 2002; Van Boven et al., 2013); the subjects in these other studies still may have predicted the given entity to be even worse (e.g., cheap wine may taste bad, but not as bad as people think it will). Regardless, these distinct literatures reveal the need for a better understanding of when negative expectations do versus do not spoil actual enjoyment.

There also remain fruitful extensions of the basic effect. For example, we focused on within-the-moment consumption, but there are other temporal dynamics to consider. Although people should presumably look back on many leisure-before-work experiences as time well spent (to the extent that they enjoyed themselves without disrupting subsequent work), retrospection of hedonic states draws on the same distanced and dulled inputs as prospection (O'Brien, Ellsworth, & Schwarz, 2012; Robinson & Clore, 2002). People may (mistakenly) regret these experiences later on because they also fail to simulate the full extent of past pleasure, which suggests an uphill battle for long-term debiasing.

Finally, it is worth emphasizing the everyday impact of an underappreciation for leisure before work. There is

always more work to do, it seems, and opportunities for leisure will inevitably arise before one feels like work is officially done. Consequently, people may intuitively push these opportunities to the future—from clerks and cashiers trying to maximize breaks and days off to entrepreneurs, creatives, and academics who craft entire schedules. The current studies reveal that at least some of these decisions may be made with the right intention but for the wrong reason. People who avoid leisure because they think the experience will be spoiled by work left undone may too often wear themselves ragged, laboring toward an ultimate payoff they could have enjoyed just as much at the start.

## Action Editor

Timothy J. Pleskac served as action editor for this article.

## Author Contributions

E. O'Brien developed the concept. Both authors contributed to the methodological designs. Testing and data collection were led by E. Roney, and E. O'Brien performed all subsequent statistical analyses. E. O'Brien wrote the manuscript, and E. Roney provided critical revisions. Both authors approved the final version of the manuscript for submission.

## Acknowledgments

We thank Jaewon Yoon, Miguel Ortega, Alex Kristal, and Chicago Booth's Center for Decision Research for helping to collect much of the data and Boaz Keysar, Anuj Shah, and Linda Hagen for especially helpful feedback.

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## Funding

This work was supported by the Willard Graham Faculty Research Award from the Booth School of Business.

## Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617701749

## Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/xmnyq. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617701749. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at https://www.psychologicalscience.org/publications/badges.

## Notes

1. For the Fixed-Labor task analyses, see the Supplemental Material. There was only a main effect of role: Predictors generally underestimated enjoyment, as in the Magic-Maker task.
2. This similarity in enjoyment was echoed via regression when we treated number of midterms continuously: Experiencers were no more likely to enjoy the spa as their workload reduced, $b = 0.01$, $SE = 0.045$, $p = .904$. We could not treat workload as a continuous variable in our main analyses because we did not have the corresponding data for predictors.
3. The finding that midterms were more distracting during the spa experience when experiencers had many midterms left to complete than when they had already finished most of their midterms was echoed via regression: The more midterms that remained, the more that experiencers were distracted, $b = 1.47$, $SE = 0.553$, $p = .009$.

## References

Ariely, D., & Loewenstein, G. (2006). Heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*, 87–98.

Ayton, P., Pott, A., & Elwakili, N. (2007). Affective forecasting: Why can't people predict their emotions? *Thinking & Reasoning*, *13*, 62–80.

Baron, J. (1990). Harmful heuristics and the improvement of thinking. In D. Kuhn (Ed.), *Developmental perspectives on teaching and learning thinking skills* (pp. 28–47). Basel, Switzerland: Karger.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289.

Buechel, E. C., Zhang, J., Morewedge, C. K., & Vosgerau, J. (2014). More intense experiences, less intense forecasts: Why people overweight probability specifications in affective forecasts. *Journal of Personality and Social Psychology*, *106*, 20–36.

Campbell, T., O'Brien, E., Van Boven, L., Schwarz, N., & Ubel, P. A. (2014). Too much experience: A desensitization bias in emotional perspective taking. *Journal of Personality and Social Psychology*, *106*, 272–285.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*, 980–1008.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Csikszentmihalyi, M. (1990). *Flow*. New York, NY: Harper and Row.

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*, 175–191.

Feather, N. T. (1999). Judgments of deservingness: Studies in the psychology of justice and achievement. *Personality and Social Psychology Review*, *3*, 86–107.

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*, 1351–1354.

Gruber, J., Mauss, I. B., & Tamir, M. (2011). A dark side of happiness? How, when, and why happiness is not always good. *Perspectives on Psychological Science*, *6*, 222–233.

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

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

Hoerger, M., Quirk, S. W., Lucas, R. E., & Carr, T. H. (2010). Cognitive determinants of affective forecasting errors. *Judgment and Decision Making*, *5*, 365–373.

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

Kahneman, D., & Thaler, R. H. (2006). Anomalies: Utility maximization and experienced utility. *The Journal of Economic Perspectives*, *20*, 221–234.

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, *43*, 259–266.

Kivetz, R., & Simonson, I. (2002). Earning the right to indulge: Effort as a determinant of customer preferences toward frequency program rewards. *Journal of Marketing Research*, *39*, 155–170.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, *110*, 499–519.

Lee, L., Frederick, S., & Ariely, D. (2006). Try it, you'll like it: The influence of expectation, consumption, and revelation on preferences for beer. *Psychological Science*, *17*, 1054–1058.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*, 272–292.

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

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, *131*, 803–855.

Morewedge, C. K., Gilbert, D. T., Myrseth, K. O. R., Kassam, K. S., & Wilson, T. D. (2010). Consuming experience: Why affective forecasters overestimate comparative value. *Journal of Experimental Social Psychology*, *46*, 986–992.

Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General*, *141*, 558–569.

Novemsky, N., & Ratner, R. K. (2003). The time course and impact of consumers' erroneous beliefs about hedonic contrast effects. *Journal of Consumer Research*, *29*, 507–516.

O'Brien, E. (2015). Feeling connected to younger versus older selves: The asymmetric impact of life stage orientation. *Cognition & Emotion*, *29*, 678–686.

O'Brien, E., & Ellsworth, P. C. (2012). Saving the last for best: A positivity bias for end experiences. *Psychological Science*, *23*, 163–165.

O'Brien, E., Ellsworth, P. C., & Schwarz, N. (2012). Today's misery and yesterday's happiness: Differential effects of current life-events on perceptions of past wellbeing. *Journal of Experimental Social Psychology*, *48*, 968–972.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, USA*, *105*, 1050–1054.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.

Ross, W. T., & Simonson, I. (1991). Evaluations of pairs of experiences: A preference for happy endings. *Journal of Behavioral Decision Making*, *4*, 273–282.

Sonnentag, S. (2012). Psychological detachment from work during leisure time: The benefits of mentally disengaging from work. *Current Directions in Psychological Science*, *21*, 114–118.

Sonnentag, S., & Fritz, C. (2015). Recovery from job stress: The stressor-detachment model as an integrative framework. *Journal of Organizational Behavior*, *36*, S72–S103.

Van Boven, L., & Epley, N. (2003). The unpacking effect in evaluative judgments: When the whole is less than the sum of its parts. *Journal of Experimental Social Psychology*, *39*, 263–269.

Van Boven, L., Loewenstein, G., Dunning, D., & Nordgren, L. F. (2013). Changing places: A dual judgment model of empathy gaps in emotional perspective taking. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 48, pp. 117–171). Burlington, MA: Academic Press.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*, 821–836.

Xu, J., & Schwarz, N. (2009). Do we really need a reason to indulge? *Journal of Marketing Research*, *46*, 25–36.