

PSYCHOLOGICAL SCIENCE

Day-to-day fluctuations in cognitive precision predict the domain-general intention-behavior gap

Daniel J. Wilson and Cendri A. Hutcherson*

The idea that better cognitive functioning helps to close the gap between goals and behavior seems self-evident, yet empirical work has found unusually weak interindividual associations between cognitive task performance and real-world, goal-relevant outcomes. Here, we resolve this paradox by shifting from trait-level to state-level analysis. Leveraging a microtask design to measure daily fluctuations in cognitive function, goal setting and goal progress, mood, sleep, and motivation over a 12-week intensive longitudinal study of university students ($N = 184$, time points = 9248), we show that within-person upswings in domain-general cognitive processing precision precede and predict same-day self-reported goal setting and achievement across both academic and nonacademic domains, even controlling for other factors. A one-standard-deviation change in cognitive precision had an effect statistically equivalent to ~40 min of work, with similar or larger predictive effects compared to fluctuations in mood/motivation and no moderation by trait-level self-control or conscientiousness. Our work addresses longstanding controversies and highlights the power of intraindividual analysis to reveal relationships missed by cross-sectional approaches.

INTRODUCTION

Humans travel the path linking goals, intentions, and outcomes every day, but our pace is often frustratingly fickle. Why? What differs between the days we achieve what we intend, and the days we fall short? Although much of the day-to-day variation in goal achievement undoubtedly stems from external obstacles, a substantial proportion is likely due to internal psychological factors. However, despite considerable empirical work on the predictors of goal achievement across individuals [e.g., (1–5)], we still have an incomplete understanding of how fluctuations in intraindividual psychological processes could shrink or grow the intention-behavior gap (I-B Gap) (6). Here, we focus on a particularly controversial issue: Do day-to-day variations in cognitive states related to information processing, attention, working memory, or inhibitory control predict changes in the effectiveness of daily goal pursuit in a domain-general way, across a broad spectrum of goals? If so, which cognitive processes matter, and how important are they relative to other predictors? Intuitively, the answer seems obvious. Because goal pursuit often requires us to ignore distractions, manage conflicting temptations, and occasionally push through unpleasant or effortful tasks for longer-term rewards (2, 7, 8), then, when processes related to cognitive control operate at their peak, people should be better at achieving their intentions and goals (9). However, the considerable efforts to document such relationships have, to date, generated more controversy than clarity (5, 10–13).

Part of the issue is that much of the research on the hypothesized benefits of optimal cognitive functioning has focused on differences across individuals rather than dynamics within individuals. While more recent computational approaches could reveal clearer links (14, 15), cross-sectional research is largely characterized by mixed and often disappointing results in healthy samples (5, 16–25). However, it is unlikely that approaches examining trait-level differences between individuals have much to say about whether day-to-day changes in the operation of specific cognitive factors predict variability in day-to-day

goal achievement. Inter- and intraindividual associations can be statistically distinct (26–28), and empirical evidence suggests that differences in cognitive task performance within versus between individuals reflect different cognitive processes and sources of variability (29). Thus, daily cognitive function could show a strong relationship to daily goal accomplishment, even in the absence of any cross-sectional association (or vice versa).

Unfortunately, an extensive body of research focused more directly on the relationship between state fluctuations in cognitive function and goal-directed behavior is similarly muddied. For example, tests of the “ego depletion” hypothesis (9) often attempt to manipulate state cognitive control and then demonstrate subsequent declines in self-control. However, these tests rarely directly measure the declines in cognitive control that are postulated to mediate the effect (11). They also frequently fail to replicate (10, 12, 30), and even the most promising studies largely focused on behavioral outcomes in laboratory tasks (31, 32) rather than real-world goal-directed decision making.

The handful of previous studies that directly measure longitudinal variation in cognitive task performance and its association with real-world goal-related behavior report mixed and often null findings (33–36), likely due in part to methodological limitations. These studies generally follow small sample sizes for a limited duration (e.g., 1 to 3 weeks); measure performance on only a single cognitive task (e.g., the Go/No-Go task); focus on a single behavioral domain (e.g., food or alcohol consumption); rarely measure in real time the changing goals people hold; and generally fail to measure or control for important alternative explanations, such as concomitant changes in affective and motivational processes (13, 37, 38). It thus remains unclear if fluctuations in cognitive function directly affect the effectiveness of daily goal pursuit, which functions matter, how much, and for which goals. Answering these questions is of vital importance, not only for resolving major theoretical controversies in psychology, but also for directing resources toward the most fruitful avenues of exploration for boosting goal achievement.

To address limitations of past research, we explore the feasibility of, and insights derived from, a microtask method for studying cognition-dependent goal achievement, both within and across

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Downloaded from <https://www.science.org> on April 29, 2026

Department of Psychology, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON M1C 1A4, Canada.

*Corresponding author. Email: c.hutcherson@utoronto.ca

individuals. Using a 12-week intensive longitudinal design (Fig. 1A, top), we provide evidence for the existence of a domain-general cognitive state variable that we refer to as mental sharpness (see Results below for justification of terminology). We argue that we can obtain a meaningful daily estimate of this state using six brief (1 to 2 min) cognitive tasks, by showing that our metric covaries with fluctuations in daily mood, motivation, and behavior. Last, we examine whether and how this variable relates to daily goal pursuit and achievement within and across individuals and compare the magnitude of effects to other important self-reported behavioral, motivational, and affective predictors. Our results support the idea that day-to-day fluctuation in mental sharpness has a pronounced, domain-general association with the daily gap between intentions and behavior and raise previously unexplored avenues for research in the quest to improve daily goal achievement, and through it, well-being (39).

RESULTS
Overview

Our analysis proceeds in five steps. First, we describe how we measured daily cognitive function, exploring the psychometric properties and behavioral/psychological correlates of our measure. Second, we describe how we measured real-world goal achievement. Third, we directly test whether and how our measure of daily cognitive function correlates with goal achievement, confirming that this correlation is not mediated by other noncognitive factors, and comparing the strength of its effect to other predictors. Fourth, we ask whether a parallel measure of trait-level cognitive performance correlates with trait-level goal achievement. Last, we engage in a multiverse analysis to illustrate the robustness of effects as well as their moderation by specific methodological and analytical choices.

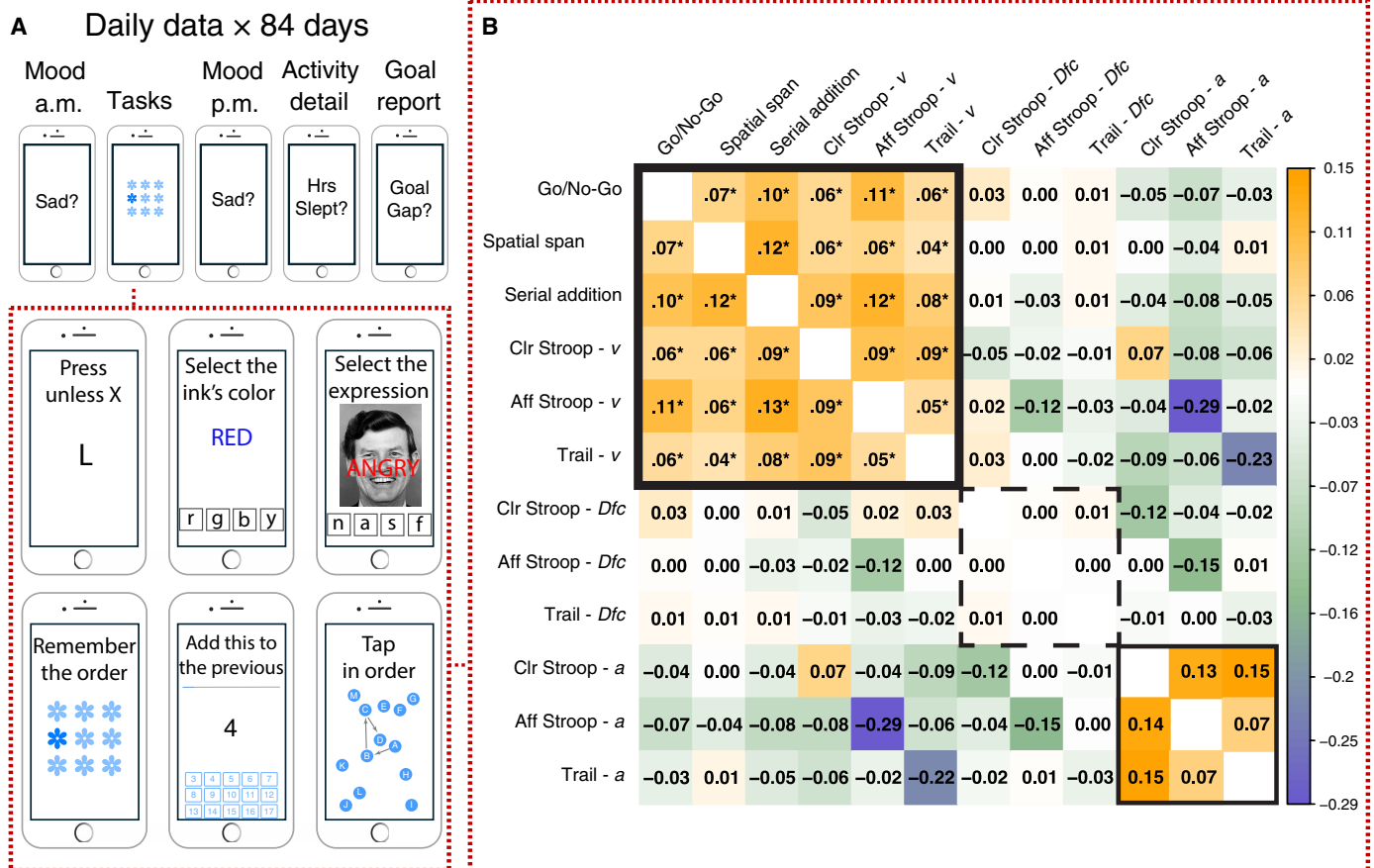


Fig. 1. Measurement of daily behaviors, goal achievement and cognitive function. (A) Top: Participants completed up to 84 days of tasks and survey questions assessing mood/subjective state, cognitive performance, details of daily activities (e.g., hours of sleep and hours of work), subjective evaluations of daily goal achievement, and setting of goals for the next day. Bottom: Cognitive performance was measured with six 1- to 2-min versions of standard cognitive tasks (Go/No-Go, color (Clr) and affective (Aff) Stroop tasks, spatial span, paced serial addition, and trail making) each day. (B) Day-to-day pairwise associations between task parameters. We focused on 12 different daily task performance parameters: Go/No-Go d' , spatial working memory span, serial addition % correct, DDM parameters related to overall drift rate v (representing differentiation of evidence for correct versus incorrect response), congruency/difficulty effects on the drift rate (Dfc), and threshold a , estimated separately for each of the Stroop and trail-making tasks. Data for each parameter were detrended and normalized within participants and then subjected to pairwise correlations in mixed effects regression models. $*P < 0.05$, Bonferroni-corrected for multiple comparisons. This revealed evidence for a common factor affecting task performance in all cognitive measures, largely related to the precision of cognitive evidence (i.e., overall drift rate in several tasks, d' in the Go/No-Go, etc.), as well as covariation of the response threshold estimated from DDMs of three tasks. We found little correlation in cross-task difficulty effects.

Measuring daily cognitive function and its properties

To measure daily fluctuations in cognition, we selected six popular cognitive tasks thought to tap different cognitive constructs and requiring judgments across different modalities (Fig. 1A, bottom; see Materials and Methods and the Supplementary Text for details and justification). These tasks included a spatial span task (40), a Go/No-Go task (41), a paced serial addition task (PSAT) (42), a trail-making task (43), a color-word cognitive conflict Stroop task (44), and a face-word affective conflict Stroop task (45). Daily performance was assessed using (i) the sum of correct recall on sequences of increasing length in the span task, (ii) mean accuracy in the PSAT, (iii) d' , a measure of discrimination accuracy in the Go/No-Go (46), and (iv) 9-day-level parameters (three per task) derived from a hierarchical drift diffusion model [DDM; (47)] of the trail-making and two Stroop tasks. DDM parameters specified signal precision (i.e., overall drift rate ν toward correct responses), the effect of congruence/difficulty on drift rate (νDfc), and response threshold (a), estimated separately for each individual and task, with day nested within participant. Simulation exercises and model convergence statistics suggested that, despite small numbers of daily trials, we can obtain stable, meaningful daily estimates of these parameters (see the Supplementary Text, note S1). We observed evidence for, and thus removed, effects of practice in all tasks (see Materials and Methods for details).

We then performed an exploratory analysis looking for evidence of a meaningful domain-general cognitive state variable (or variables) influencing behavior across some or all task parameters completed on a given day. To test this, we looked for evidence of pairwise daily covariation among the normalized, detrended residuals of task parameters, using a series of mixed-effects models predicting one parameter from another. We found evidence for such covariation, with significant positive cross-task correlations (all $P < 0.05$, Bonferroni-corrected for 132 comparisons) between spatial span, PSAT accuracy, Go/No-Go d' , and overall drift rates in trail making, cognitive, and affective Stroop tasks (Fig. 1B). We also observed cross-task covariation in response thresholds, although task-specific thresholds did not consistently correlate with other cross-task parameters. We found no detectable cross-task covariation in difficulty effects.

These results suggest that six parameters related to overall performance on different tasks jointly reflect a domain-general cognitive process, distinct from conflict resolution/difficulty effects or response caution. However, the standardized pairwise regression coefficients across parameters ranged from 0.04 to 0.13 ($M = 0.081$), suggesting substantial measurement noise in any single task. We thus took the average of these six parameters as our measure of daily cognitive function, which statistical theory suggests should have a correlation of ~ 0.6 with the true underlying state, given the magnitude of the individual correlations among items (48). Put differently, our measure captures around 34% of the variance in cognitive fluctuations, equivalent to Cronbach's α of 0.34. We accepted this level of measurement imprecision to enable measurement efficiency, keeping the completion time of our cognitive assessments to a manageable 10 min per day and allowing us to retain participants over a longer study duration. This ultimately also means that any relationship observed between our measure of cognitive function and goal achievement is almost certainly an underestimate of its true magnitude, given measurement noise.

In the remainder of this manuscript, we refer to this six-task composite score as daily mental sharpness. We use this term, rather than a term like cognitive efficiency (14) or information processing speed (15), for several reasons. First, it provides a more intuitive and generally accessible description of what we suspect our measure ultimately

reflects. Second, although our cognitive measure correlates most clearly with objectively derived computational parameters related to the strength/sharpness of signal relative to noise in information processing, it also correlates with performance metrics in tasks that are not as amenable to computational modeling and thus cannot directly be tied to evidence accumulation processes. Last, we opted for this term because our measure correlates most strongly with the subjective phenomenology of feeling sharp and focused rather than distracted while completing tasks (see below).

Properties of daily mental sharpness

To bolster the conclusion that despite measurement noise we can estimate a domain-general cognitive state on a day-to-day basis with sufficient accuracy for statistical inferences, we show that fluctuations in our mental sharpness composite are not random, but have clear temporal structure and covary with theoretically sensible factors related to behavior, mood, and motivation (see Materials and Methods for model details, standardized β coefficients reported unless otherwise noted). First, we find that mental sharpness shows strong autocorrelation across days (Fig. 2A): Sharpness up to 3 days prior (even accounting for intervening days) predicted today's sharpness, with declining effects at increasing intervals ($\beta_{\text{Day}-1} = 0.086 \pm 0.017$, $P < 0.00001$; $\beta_{\text{Day}-2} = 0.048 \pm 0.013$, $P = 0.0003$; $\beta_{\text{Day}-3} = 0.040 \pm 0.012$, $P = 0.001$; $\beta_{\text{Day}-4} = 0.014 \pm 0.012$, $P = 0.25$). Second, we show that our measure demonstrates moderate fatigue effects: It is higher if participants reported sleeping more hours than typical the night before ($\beta_{\text{Sleep}} = 0.024 \pm 0.01$, $P = 0.02$) and declines continuously over time within the day (unstandardized $\beta_{\text{Hour}} = -0.006 \pm 0.0009$, $P < 0.00001$; Fig. 2B).

Consistent with other findings of relations between mood and cognition (49), our measure of daily mental sharpness also correlated with self-reported mood (Fig. 2D). It was lower when people had felt more depressed ($\beta_{\text{Depressed}} = -0.04 \pm 0.012$, $P = 0.002$), excited ($\beta_{\text{Excited}} = -0.04 \pm 0.013$, $P = 0.0008$), or (marginally) angry ($\beta_{\text{Anger}} = -0.022 \pm 0.012$, $P = 0.06$); was higher when they felt more generally motivated or anxious ($\beta_{\text{Motivated}} = 0.049 \pm 0.013$, $P = 0.0002$; $\beta_{\text{Anxious}} = 0.031 \pm 0.013$, $P = 0.02$); and was unrelated to feeling happier or calmer (both $P > 0.15$). The composite measure also correlated with task-related self-report (Fig. 2E). It was higher when participants endorsed items from the NASA task load index (NASA-TLX) related to performance ($\beta_{\text{Performance}} = 0.16 \pm 0.014$, $P < 0.00001$) and effort ($\beta_{\text{Effort}} = 0.055 \pm 0.013$, $P < 0.00001$) and correlated negatively with frustration/negative affect ($\beta_{\text{Frustration}} = -0.033 \pm 0.013$, $P = 0.02$). Composite scores were unrelated to finding the tasks mentally demanding or temporally rushed (both $P > 0.2$). Mental sharpness also correlated with two additional items tapping task-related engagement, being higher on days when participants reported before the tasks that they felt more motivated to perform well ($\beta_{\text{TaskMotivation}} = 0.029 \pm 0.013$, $P = 0.02$) and lower when they felt more distracted ($\beta_{\text{Distraction}} = -0.11 \pm 0.013$, $P < 0.00001$), reported after task completion.

We also found evidence of effects related to working harder than average, especially over extended periods of time (Fig. 2C): While mental sharpness was higher on days that participants reported having worked longer than average ($\beta_{\text{Today's Work}} = 0.033 \pm 0.01$, $P = 0.006$) and was not affected by working longer than average the previous day ($\beta_{\text{Previous Day's Work}} = -0.006 \pm 0.013$, $P = 0.67$), it was significantly lower if the past week had involved longer-than-average work days ($\beta_{\text{Previous Week's Work}} = -0.037 \pm 0.014$, $P = 0.007$). The effect of a long work week persisted ($\beta_{\text{Previous Week's Work}} = -0.036 \pm 0.013$, $P = 0.007$)

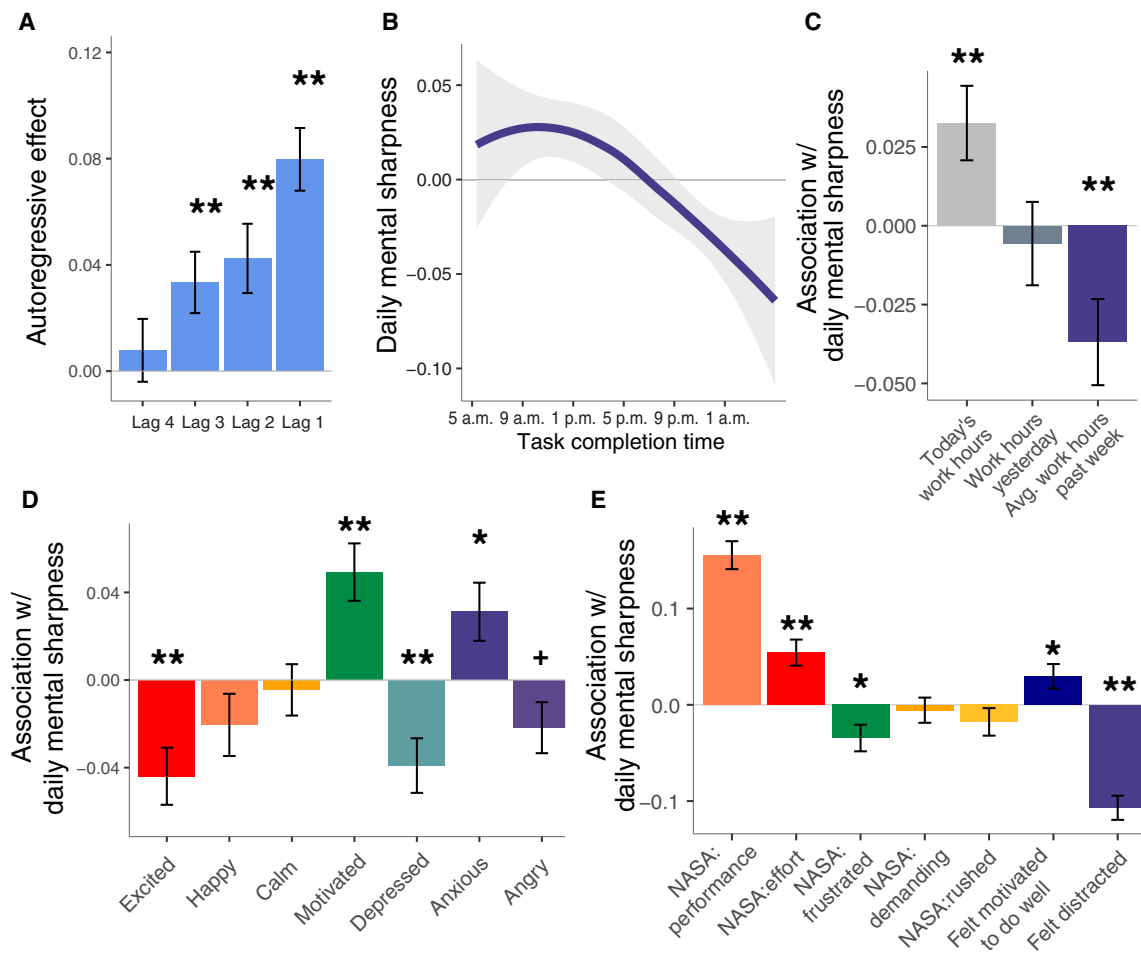


Fig. 2. Properties of daily mental sharpness. Our CC of mental sharpness showed (A) strong autoregressive effects across days, correlating with sharpness up to 3 days prior; (B) declines over time within a day; (C) a tendency to be lower after reporting having worked more hours than average over the past week, but not after a single day; (D) significant relationships with daily mood; and (E) strong relationships with self-reported task load (NASA-TLX items), as well as motivation to perform well (reported prior to the tasks) and distraction (reported after). Bars represent the unique, standardized effect of a variable estimated from a mixed-effect regression including all variables displayed within a given panel. Error bars represent standard error of the β coefficient. + $P < 0.1$; * $P < 0.05$; ** $P < 0.005$.

even after controlling for self-reported daily affective motivation, task-specific motivation, distraction, and prior sleep, suggesting that it was not simply due to opportunity costs or labor/leisure-style shifts in desire to work (13, 38).

Characteristics of goal completion

Before turning to the relationship between mental sharpness and goal pursuit, we next describe how we measured goal achievement, as well as its characteristics (see Fig. 3). As with cognitive state, we used a combination of measures, largely focused on the gap between intentions and outcome. First, every evening, participants briefly described two goals they wanted to complete the next day (sample responses: “Finish bio homework” and “Get car fixed once n for all”). Participants also answered questions about each goal, including anticipated effort, importance, internal and external motivation, confidence in completion, and consequences for failure. The next day, we reminded participants of their two freely described goals and asked them to rate what percentage (0 to 100%) of each goal they had completed.

We also obtained more general measures of goal completion, by asking participants to indicate their subjective estimate of the overall

gap between the things they had intended to do and what they actually did, on a scale from 0 (no gap) to 100 (got nothing done of what they wanted). Participants also indicated how productive and focused they had felt, as well as how much they had procrastinated.

Last, participants answered binary yes/no questions about whether they had goals for the next day in specific domains (e.g., work, minutes of exercise, time spent in leisure activities, etc.). If participants indicated having a goal, they were prompted to quantify this goal (e.g., amount of time desired). We then compared this quantity to the quantity they indicated completing the next day (without reminding them of their previous intention). Participants endorsed domain-specific goals relatively rarely, except in the case of work hours. We thus report analyses only for the I-B Gap between desired and actual work hours, computed on days for which participants had a goal the previous evening (~30% of days).

On average, participants reported both considerable success, as well as notable failures, in goal completion. They completed on average $61.5 \pm 14\%$ of their two freely described specific goals, had a self-estimated overall gap of about $47.6 \pm 14.7\%$, and reported working 1.13 ± 1.76 hours less per day than intended. Participants’

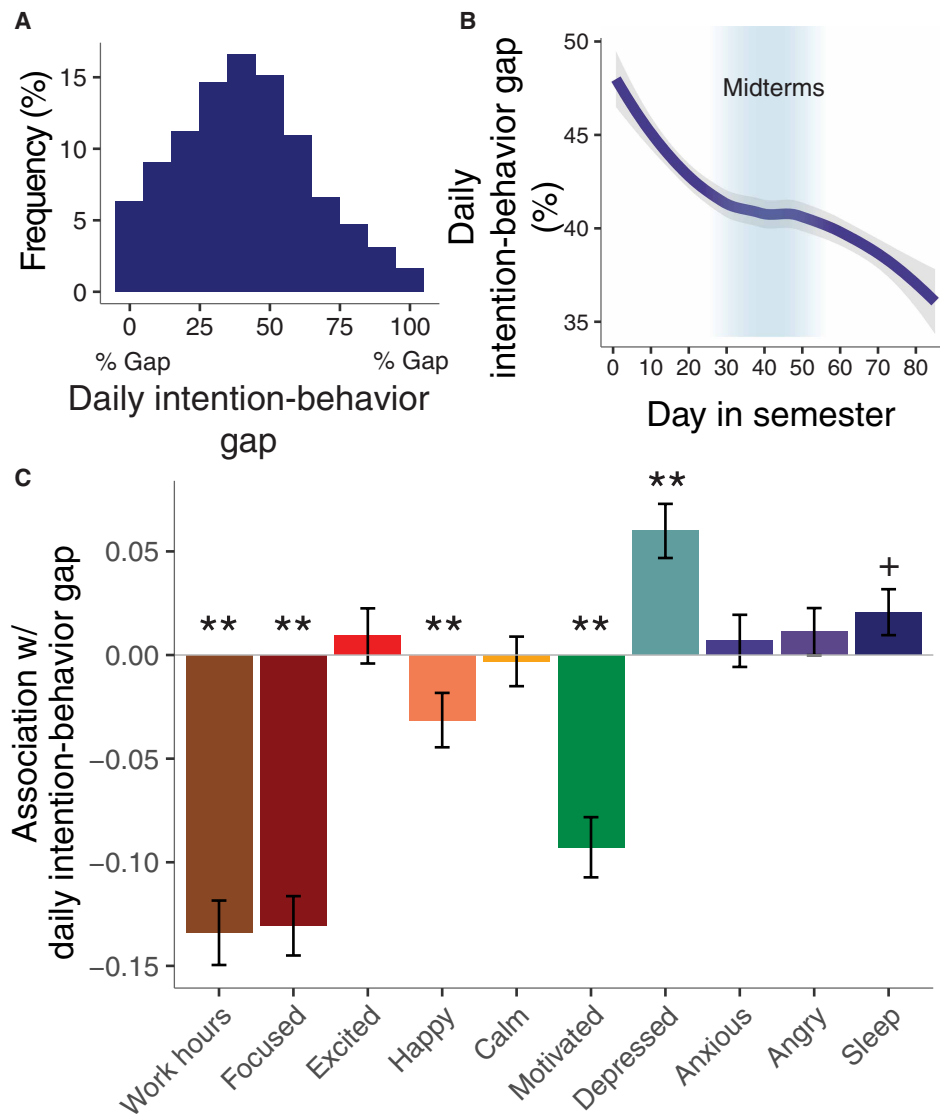


Fig. 3. Exploring the daily I-B Gap. The daily gap, measured as a percentage of desired achievement (0 = no gap, 100 = total gap) and averaged across self-reported gaps for two specific goals and a question assessing the overall gap on a given day, showed (A) a broad distribution across days; (B) declines on average over the course of a semester; and (C) strong associations with behavior, mood, motivation, and subjective focus. Bars represent the unique effect of each variable, standardized within participant, estimated from a mixed-effect regression including all variables displayed in the panel. Error bars represent standard error of the estimated regression coefficient. * $P < 0.05$; ** $P < 0.005$.

self-reported gaps in both specific goals and overall (all $P < 0.03$) decreased over the semester (Fig. 3B), although this trend did not extend to the work hour gap ($P = 0.44$). Because the source of these long-term trends is unclear, we conducted analyses on the relationship between cognition and goal achievement both with and without linear detrending of goal achievement measures. This had little effect on conclusions (Supplementary Text, note S2).

Last, we explored how goal achievement related to behavior, mood, and motivation (Fig. 3C). Because individual goal measures correlated with each other (all $P < 0.0001$), we created a composite measure of the I-B Gap using normalized, detrended versions of the three subjectively reported goal completion items (i.e., the overall gap and the two specific goal reports, reverse-scored). We refer to this composite measure as the I-B gap. We then assessed daily covariation in the I-B gap and hours worked, hours slept, mood, motivation

and focus (see Materials and Methods for model details). Working longer than average hours was the strongest single correlate of the I-B gap ($\beta = -0.134 \pm 0.015$, $P < 0.00001$), followed by feeling focused ($\beta = -0.13 \pm 0.014$, $P < 0.00001$) and feeling motivated during the day ($\beta = -0.093 \pm 0.015$, $P < 0.00001$). Feeling happy ($\beta = -0.031 \pm 0.013$, $P = 0.02$) was negatively associated with the gap, while feeling depressed correlated positively ($\beta = 0.060 \pm 0.013$, $P < 0.00001$). The I-B gap was marginally higher after sleeping more than average ($\beta = 0.020 \pm 0.011$, $P = 0.06$) and unrelated to feeling excited, calm, anxious, or angry, after accounting for other factors.

The relationship between state fluctuations in mental sharpness and the I-B Gap

Does daily cognitive function relate to daily goal achievement? A mixed-effects regression predicting daily fluctuations in the composite

I-B gap from daily fluctuations in mental sharpness suggested that it does ($\beta = -0.047 \pm 0.011$, $P < 0.0001$; Fig. 4A). Supplementary analyses also confirmed that this relationship was not driven by any single item from the composite gap measure but was consistently observed across different goal-related items separately (Supplementary Text, note S2). It was also not affected by accounting for cyclic patterns related to the day of the week (Supplementary Text, note S3). Cognitive composites (CCs) related to response caution (i.e., the average of parameters related to the estimated decision threshold a from tasks modeled by the DDM) and conflict resolution (i.e., the average of parameters related to the effects of difficulty on DDM drift rates $vDfc$) did not consistently correlate with goal achievement (Supplementary Text, note S2).

We next asked whether the association between mental sharpness and goal achievement was mediated by other variables. Sharpness remained strongly and independently associated with the I-B gap ($\beta = -0.034 \pm 0.011$, $P = 0.001$) even after controlling for other variables that correlated with both mental sharpness and the composite gap, including total work hours, motivation, self-reported focus, depression, and sleep. Moreover, the relationship between mental sharpness and the I-B gap was not simply due to people performing more poorly on cognitive tasks in the evening after they had experienced repeated goal failures during the day: The relationship remained

significant when restricting analysis to the 24% of days where cognitive tasks were completed before noon, a minimum of 8 hours prior to the report of goal completion ($\beta = -0.059 \pm 0.025$, $P = 0.02$), suggesting that higher morning levels of mental sharpness preceded and predicted goal-directed behavior throughout the remainder of the day.

Comparing mental sharpness to other behavioral and motivational factors

Although these results suggest a significant relationship between daily mental sharpness and goal achievement, effect sizes were generally small. We thus compared the magnitude of these effects to other predictors of the I-B gap. We focused on three variables, including the best single-item behavioral predictor of goal achievement (hours worked in a day, reported in the evening), as well as two composite predictors: a two-item composite of “subjective readiness” combining self-reported motivation and focus (both of which correlated highly with goal achievement), and a three-item composite of negative mood, combining self-reported depression, anxiety, and anger. Because items making up these composites were reported twice, once in the morning as concurrent affect and once in the evening as a retrospective average over the day, we computed separate

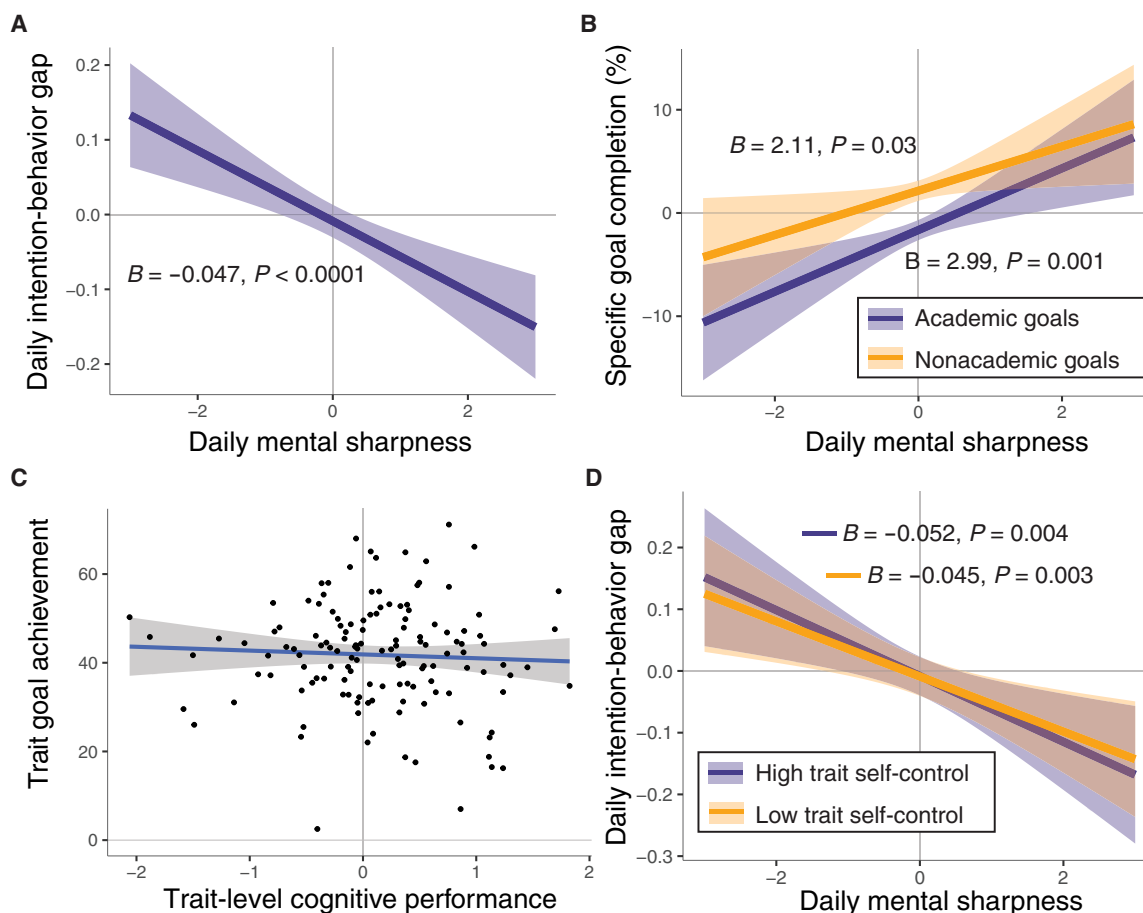


Fig. 4. Mental sharpness and the I-B Gap. (A) We observed lower daily I-B Gaps on days with higher daily mental sharpness. (B) Higher achievement of specific goals when daily mental sharpness was high occurred for both academic and nonacademic goals, with no significant difference between them. (C) Interindividual differences in trait-level (i.e., average) goal achievement were unrelated to interindividual differences in trait-level cognitive performance. (D) Individual differences in trait self-control (displayed here as a median split) did not moderate the relation between daily mental sharpness and daily goal achievement.

composites for both time points. To compare effects, we normalized each measure within the participant and used it, along with mental sharpness, to predict the normalized composite I-B gap, resulting in standardized β coefficients for comparison.

For work hours, this analysis revealed that a 1-SD change in mental sharpness had ~26% of the impact of a 1-SD change in hours worked. Given that 1 SD in work equaled 2.59 hours on average across participants, this suggests that being 1 SD or more above one's mean in mental sharpness (which would be expected on 16% of days, or approximately once a week) predicts a change in the gap that is statistically equivalent to working an additional 40 min in a day. Conversely, a 1-SD decline in sharpness is roughly like losing 40 min of work that day. We emphasize, however, that this estimate should be considered an approximation, with the precise equivalence likely dependent on a variety of factors including the nature and time commitment required of the goals.

We approached the comparison to subjective readiness and negative mood in two ways. First, we compared daily mental sharpness to subjective ratings made at day's end. This revealed that a 1-SD increase in sharpness had ~16% of the explanatory power of a 1-SD increase in average subjective readiness (Cronbach's $\alpha = 0.66$) throughout the day and ~73% that of a 1-SD change in negative mood (Cronbach's $\alpha = 0.60$). However, these comparisons likely underestimate the predictive power of mental sharpness, since evening readiness and mood reports were assessed retrospectively as estimates over the whole day and could be biased or informed by assessments of the gap itself, whereas cognitive tasks were assessed at a single point in time and often in the morning, prior to most of the day's activities. To provide a fairer comparison, we examined the predictive power of mental sharpness and morning self-reports of subjective readiness (Cronbach's $\alpha = 0.61$) and negative mood (Cronbach's $\alpha = 0.49$), restricted to days when tasks and self-reported metrics were both completed prior to 12:30 p.m. (the cutoff for morning reports of subjective experience). Compared to morning reports of subjective readiness, daily mental sharpness had greater (1.33 \times) predictive power, although both independently and significantly predicted the gap ($\beta_{\text{Readiness}} = -0.046$, $P = 0.04$; $\beta_{\text{Sharpness}} = -0.061$, $P = 0.01$). A composite of morning negative mood failed to predict the I-B gap later in the day, and mental sharpness had a much larger (~13.5 \times) effect ($\beta_{\text{Mood}} = 0.005$, $P = 0.83$; $\beta_{\text{Sharpness}} = -0.063$, $P = 0.008$). Thus, our CC measured at a single time point in the morning had comparable or greater predictive power for goal achievement over the rest of the day compared to self-reported motivational and affective factors measured at roughly the same time.

The influence of daily mental sharpness is not specific to cognitive goals

We next asked whether mental sharpness predicts only the completion of cognitive/academic goals in our student sample or whether it might relate to more general goal achievement. To answer this question, we used a large language model (LLM) to categorize each of the two freely described daily goals as academic (i.e., goals pertaining to schoolwork and learning, collectively representing 46% of the sample) or nonacademic goals (Cohen's $\kappa = 0.86$ with human ratings; see the Supplementary Text, note S4 for details). We then estimated self-reported achievement of each specific daily goal from three variables: daily mental sharpness, a binary variable indicating goal category (academic/nonacademic), and their interaction. Although we observed a main effect of goal category, indicating that academic goals were harder to achieve ($\beta = -4.73 \pm 0.715$, $P < 0.00001$), there

was no significant difference in the effect of mental sharpness as a function of goal type ($\beta = 1.38 \pm 1.11$, $P = 0.21$), which continued to correlate with goal achievement when considering only nonacademic goals ($\beta = 2.11 \pm 0.96$, $P = 0.03$; Fig. 4B). However, we did observe an important relationship between mental sharpness and whether goals were academic or nonacademic. On days when sharpness was high, participants were significantly more likely to set academic goals for the next day ($\beta = 0.14 \pm 0.04$, $P = 0.0001$). These results suggest a domain-general effect of mental sharpness on goal achievement, as well as a domain-specific effect on the tendency to set academic rather than nonacademic goals for the coming day.

Trait-level individual differences

Our work highlights a relationship between intraindividual fluctuations in mental sharpness and goal achievement. However, research on cross-individual differences in cognitive function suggests that it bears little relation to either self-reported self-control (17, 22, 23) or to important life outcomes (5). Although the sample size for this analysis ($N = 137$ participants with at least 28 complete days of task data) is somewhat underpowered to detect cross-individual differences, we nevertheless sought to test, as a point of comparison, whether there are interindividual relationships in addition to intraindividual ones. To assess trait-level cognitive function, we used a widely used method in longitudinal data analysis (50) to estimate individual intercepts after accounting for learning/growth (see Materials and Methods for details). These estimates were derived for each of the six cognitive tasks (e.g., Go/No-Go d' , overall drift v in the color Stroop, etc.), normalized across participants within a task, and then averaged for each individual (Cronbach's $\alpha = 0.82$). We then estimated a mixed-effects model predicting the raw composite I-B gap (i.e., the day-to-day average of the raw values of our three goal report items) from both the person-level estimate of cognitive performance (normalized at the group level for comparability to within-subject effects) and our normalized within-subject measure of daily mental sharpness. This model also included day-in-study to account for linear trends in the raw data. This model confirmed that within-person daily mental sharpness was associated with the raw daily gap ($\beta = -0.968$, $P < 0.0001$). However, between-person variation was not ($\beta = -0.437$, $P = 0.68$; Fig. 4C). We also note that not only was the estimated between-person effect less than half the strength of the within-person effect, but the higher reliability of the trait-level composite also implies considerably less attenuation of the estimate.

We next sought to determine whether trait-level cognitive performance correlated with other measures of trait-level self-reported goal achievement, such as conscientiousness (51), self-control (2), and grit (52). While inclusion of these trait-level measures in models of the raw daily I-B gap confirmed that they correlated with lower overall I-B gaps (all $\beta < -3.44$, all $P < 0.0001$), we observed little correlation between these measures and trait-level cognitive performance (all $r_{132} < 0.12$, all $P > 0.18$).

Last, we asked whether conscientiousness, self-control, or grit moderate the relationship between daily fluctuations in mental sharpness and goal achievement (Fig. 4D). This analysis tests the hypothesis that people who are better able to achieve their goals are also more resistant to fluctuations in their cognitive state. However, we found no moderating effect of any personality variable (all $P > 0.31$). We also did not find any evidence that conscientiousness, grit, or self-control buffer individuals against low cognitive states, while allowing them to take advantage of high cognitive states: The moderating effect of personality was no stronger when mental sharpness was below a person's

average than when it was above. Thus, although conscientious and self-controlled individuals more frequently achieve their goals on average, they show the same relationship between daily fluctuations in mental sharpness and goal completion as anyone else.

Multiverse analysis

The results reported above represent a hypothesis-driven but exploratory set of analyses that were not formally preregistered. We thus engaged in a multiverse analysis (53) to ensure robustness to experimenter degrees of freedom, including decisions about subject-level exclusions, task measures, task preprocessing, and computational modeling. In the primary analysis reported above, we estimated the association between the I-B gap and mental sharpness, as measured over 12 weeks by a combination of six different task parameters (including three estimated from day-level DDMs). Our analysis included removing temporal trends, excluding the first and last 2 days of the time series based on visual inspection of residuals and excluding any participant who did not have at least 28 days' worth of observations for both the I-B gap and daily mental sharpness. However, a number of these decisions represented "informed guesses" as to good design and analytical approaches, with reasonable arguments for alternatives. For example, we could have collected less data from participants (e.g., asking them to complete daily measures for 3 weeks instead of 3 months). We could have used only a subset of the cognitive tasks or used different metrics on these tasks (e.g., percent accuracy for the Stroop rather than drift rate). We could have opted not to remove practice effects or not to remove "beginning and end" effects during temporal detrending. Any of these choices could have affected the strength of our conclusions.

We thus engaged in a multiverse analysis (53), analyzing the relationship between the I-B gap and daily mental sharpness under different assumptions and approaches that varied the specification of the CC as well as details of study design and analysis (see Materials and Methods for details). The combination of these alternatives yielded 144 "alternative candidates" for analytical pipelines. Heterogeneity of findings across these alternatives might be a basis for low confidence in our findings. However, our multiverse analysis of these 144 combinations suggested unexpectedly robust and reliable effects, with 116 (80.6%) yielding a significant association ($P < 0.05$) between mental sharpness and the I-B gap (see Fig. 5A for a histogram of P values).

Nevertheless, the patterns of significance and nonsignificance (Fig. 5B) make clear that two design decisions have an important influence on the strength of effects. First and most clearly, using CCs of daily mental sharpness that rely only on traditional metrics of accuracy from conflict tasks, rather than using the DDM to estimate overall evidence precision, generally weakens results. For example, analyses using a CC that relied exclusively on traditional metrics from the Go/No-Go and two Stroop tasks (CC5 in Fig. 5) showed weaker relationships to goal achievement, with only 33% of analyses falling below the $P < 0.05$ threshold. Notably, a CC of the same three tasks but modeling two of them with the DDM (CC3 in Fig. 5) increased the frequency of observing significant relationships between mental sharpness and the I-B gap from 33 to 83% of cases. These results support other work suggesting that computational modeling of responses with the DDM may help to overcome other limitations of traditional metrics (14). Second, excluding the first few days of task data, when learning effects are most pronounced, has a notable, although subtler effect: 60.4% of analyses without any exclusions found significant effects, suggesting that there

may be early learning dynamics or noise that are not fully captured by our models, adding heterogeneity to the observations.

Last, just as interesting as the analytical choices that mattered were those that did not. Somewhat contrary to our expectations, analyzing shorter time periods (3 weeks instead of 3 months) or excluding participants with a smaller percentage of completed days generally did not alter effects when the CC measure itself was robust. This suggests that researchers interested in exploring these relationships might be able to recruit individuals for shorter periods of time and still observe notable effects. For example, if a researcher were interested in efficient data collection, our results suggest that responses could be collected from just three tasks (the Go/No-Go and two Stroop tasks, which in combination take only 5 min per day to complete) and for as little as 3 weeks. When modeling Stroop tasks with the DDM, dropping the first week to avoid practice effects, and excluding low-frequency responders, our results suggest that such a design could still produce robust effects at a level of $P < 0.001$. We hope that such observations will prove useful to researchers and enable more studies on the role of daily cognitive function in shaping behavior and goal achievement, even with limited resources.

DISCUSSION

Attempts to understand the psychological processes that promote goal achievement have resulted in contentious debates about the relative importance of cognitive functions related to information processing, attention, and cognitive control (5, 11, 13, 14, 17, 22, 23, 30, 54–57). By combining brief mobile phone–based assays of cognitive performance with computational modeling and daily assessments of real-world goal pursuit over a 12-week period, we contribute to this debate, providing the most robust evidence to date that state variability in what we term mental sharpness, a factor associated with the precision or clarity of information processing (14, 15), is associated with subjective daily goal progress across a variety of goals. This association persists even after accounting for other important predictors of goal pursuit, including sleep, work hours, mood, and motivation. Crucially, the magnitude of this effect is nontrivial: A 1-SD boost in mental sharpness is associated with changes in goal achievement comparable to an additional 40 min of work, and the power of this cognitive factor to predict how the day's goal achievement will unfold is similar to or even exceeds self-reported measures of motivational and mood states. Despite these robust intra-individual relationships, trait-level cognitive performance and goal achievement were unrelated. Our findings may help to reconcile competing theories and observations about the predictors of goal achievement (6) and point to previously unidentified avenues for exploration.

Our use of a variety of cognitive tasks to measure different aspects of daily cognitive function helps to illuminate both the domain general nature of daily fluctuations in cognitive performance, as well as which aspects of cognition may matter most. We find that our metric of mental sharpness, conceptually related most clearly to the clarity of information processing and anchored in measures of signal precision like DDM drift rate, outperformed indices of conflict-specific effects or response caution. One possibility for why we observe such effects is that many real-world goals hinge more on the demand for sustained, high-fidelity processing of task-relevant information in noisy environments, rather than the ability to inhibit or resolve conflicting responses per se. This perspective aligns with theoretical work suggesting that cognitive efficiency, the ratio of task-relevant information to internal and external noise, may be a central currency of self-regulation (14),

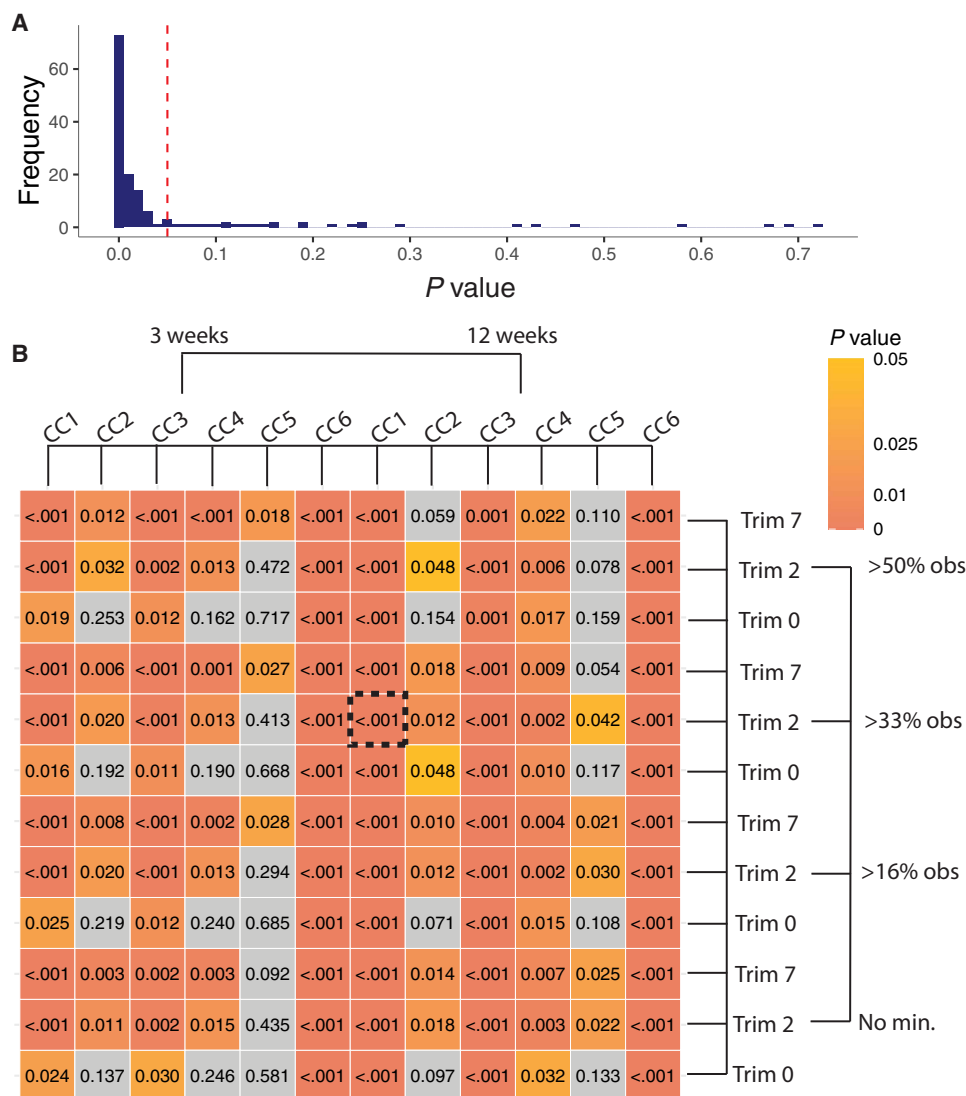


Fig. 5. Assessing robustness via a multiverse analysis. We conducted 144 different replications of our central finding examining the association between daily mental sharpness and the I-B gap. **(A)** The distribution of *P* values from these analyses suggested a robust effect, with the majority of analytical choices (80.6%) resulting in significant *P* values. **(B)** Examining patterns of significance and nonsignificance with different combinations of methodological and analytical choices yields several insights into the choices that affect the strength of the relationship. Rows: two levels of choices involving whether to exclude participants who fail to meet a minimum (min.) threshold for total observations (obs) and whether to trim/exclude the first few and last few days of tasks (trim of 0, 2, or 7 days). Columns: two levels of choices involving collecting only 3 weeks of data or 3 months, as well as using one of six different variants of a CC mental sharpness measure. CC1: the CC measure reported in the primary analysis. CC2: a measure that included only spatial working memory and serial addition tasks. CC3: a measure that included only the Go/No-Go and overall drift rates from the color and affective Stroop tasks. CC4: a measure that included all tasks, but used overall accuracy instead of overall drift rate *v* for the two Stroop and trail-making tasks. CC5: a measure that included only the Go/No-Go and accuracy measures from the color and affective Stroop task. CC6: measure using all six of the primary analysis task parameters, as in CC1, but with no detrending to remove practice effects. Dashed square indicates the specific combination of choices leading to the results reported in the primary analysis.

and that cognitive control may be largely irrelevant for real-world goal pursuit (5, 17, 23, 55). However, it is also possible that metrics related to daily inhibitory control or response caution are computed less reliably or require more trials to cleanly measure (58) compared to overall signal precision, thus obscuring their predictive associations. Our results thus speak more to the importance of the efficiency of cognitive processing, rather than speaking definitively against the importance of other cognitive factors.

Our findings have several exciting practical implications. Because mental sharpness is malleable on a day-to-day basis and was tied to

modifiable factors like sleep quantity, prior workload, and emotional state, it represents a promising target for interventions. However, our work highlights the unexpected nature of some of these associations: Mental sharpness decreased not only when feeling depressed but also when feeling excited, and it increased when feeling anxious. These findings defy an easy separation by valence or arousal and suggest that a focus on specific combinations of emotional states may foster greater mental sharpness and perhaps goal achievement.

Our work also highlights important practical trade-offs in navigating goal pursuit over extended periods of time. For example, in

our sample of university students, working longer hours predicted a lower I-B gap. An extended period of above-average work hours also predicted lower cognitive function, although a single long workday did not. These dynamics corroborate models positing extended timescales of fatigue and recovery (32, 59), and caution against interventions that push individuals to chronically high workloads without adequate recovery. Future work could investigate whether providing participants knowledge about their momentary cognitive state or its decline over time could help them to make more adaptive choices about when to work, as well as when and how to rest.

Future work will also be needed to address some of the most important limitations of the present investigation. First and foremost, our microtask approach (which emphasized measurement efficiency over precision) comes with considerable measurement uncertainty, curtailing its practical utility. Our approach as currently implemented is likely too noisy to be used for within-person research to reliably inform a person about their cognitive state, or to trigger just-in-time interventions (60), for instance. Although the strength and consistency of our findings are perhaps more remarkable given such noise, future work should explore the optimal combination of tasks and analytic approaches for efficient-yet-reliable measurement of this important construct. Second, our conclusions relied on self-reported goal achievement, which can diverge in important ways from objective measures of behavior and outcomes (61). Future work could address this dichotomy by combining our microtask approach with passive digital trace data (62) tracking objective behaviors related to daily work, physical activity, and social engagement. Third, our participants were young, healthy adults who showed relatively high conscientiousness in completing daily tasks and self-reported measures. This might restrict our ability to detect relationships between interindividual differences in cognition and goal achievement by eliminating the lower end of the ranges for both measures. Moreover, cognition-behavior links might differ in children, working-age and older adults, clinical groups, or groups with markedly different goals and daily rhythms, a possibility that could be explored through recruiting more diverse samples. Last, the observational nature of our work precludes clear causal inferences, and we urge caution in making such claims here. However, we do think this work justifies the potential utility of experimental work exploring behavioral or pharmacological interventions that increase mental sharpness, which could yield theoretical and practical insights into the utility of cognition as a target for fostering goal achievement.

Our findings thus open a raft of questions, reviving the intuition that cognitive functions matter for goal pursuit, but also forcing a re-examination of when and how they matter. Rather than a trait-level predictor of success, mental sharpness emerges as a fluid, measurable state that waxes and wanes with sleep, workload and mood and, in turn, predicts differences in the daily gap between what people intend and what they achieve. Recognizing and leveraging these fluctuations may foster previously unexplored interventions, technologies, and organizational practices that help people pursue what they care about, one day at a time.

MATERIALS AND METHODS

Participants

In total, we recruited 184 undergraduate students (mean age of 18.7 ± 1.3 SD; 36 male, 124 female, 3 nonbinary, and 21 declined to answer; 5.9% White, 54% Asian or South Asian, 10.8% Black, 16.8% other, and 11.4% no answer) using fliers distributed outside an

Introductory Psychology class at the University of Toronto Scarborough. Participants were required to be above the age of 18 and to own an Apple iPhone. Data collection took place in two cohorts over two 12-week (84-day) periods, with the first cohort collected from 27 September 2022 to 19 December 2022 (inclusive), and the second cohort collected from 30 January 2023 to 23 April 2023 (inclusive). Participants were compensated \$387 on average. Because our goal was to study daily dynamics of the within-person relationship between cognitive states and goal achievement, we included in the primary analyses only participants with at least 28 days on which they completed both the evening survey item assessing their overall goal completion and any subset of the six cognitive tasks on that day ($N = 137$; mean age of 18.8 ± 1.4 ; 27 male, 103 female, 3 nonbinary, and 4 declined to answer; 7.3% White, 59.9% Asian or South Asian, 12.4% Black, 16.8% other, and 2.9% no answer). The study was approved by the Research Ethics Board at the University of Toronto.

Procedure

Overview

Participation took place over the course of one university semester (3 months). Upon recruitment, participants received all study details, provided informed consent, were loaned an Apple Watch, and received help downloading and setting up MyDataHelps, a cross-platform mobile phone application developed by CareEvolution, Inc. that enables delivery of surveys and tasks to users. At this time, participants also received instructions for and completed one round of each daily phone-based task and were allowed to ask any questions they had about tasks, surveys, or participation. Following onboarding, participants had up to 1 week to complete a battery of 35 self-reported individual difference measures and 22 task-based cognitive measures (Supplementary Text). These tasks and surveys were largely derived from a battery of measures used to develop an ontology of self-regulation (5). Participants also began completing the daily task and self-reported measures, continuing for up to 84 days. To promote better motivation, we developed a custom dashboard within the MyDataHelps platform, providing participants with information about their earnings to-date, expected earnings at the end of the study, receipt of bonus performance rewards, trends on task and survey performance, and overall patterns of daily activity. At the end of this period, participants had another week to complete a second round of the same 35 individual difference surveys and 23 task-based measures completed at the beginning of the study. Last, participants returned the loaner watch, were debriefed, and paid according to their total earnings (mean = \$387, median = \$432.50).

Cognitive microtasks

Each day, participants completed a set of 10 (Cohort 1) or 11 (Cohort 2) “microtasks” (1- to 2-min-long versions of popular tasks for assessing cognition, motivation, and decision-making). These tasks became available at 5:15 a.m. each day, with participants receiving a push notification to complete them. Participants were encouraged to complete tasks first thing in the morning, but were allowed to complete them at any point during the day up until 4:45 a.m. (i.e., 23.5 hours after release). Participants thus received a second push notification at 10:15 a.m. if they had not yet completed the daily tasks. Time of final task completion each day followed a bimodal distribution with peaks at ~10:00 a.m. and ~11:00 p.m. (fig. S1). Participants completed the tasks before noon or after 8:00 p.m. on 29 and 49% of days, respectively. Here, we provide details for the six tasks comprising our measure of daily cognitive function. In

the Supplementary Text, we also provide details regarding the remaining tasks.

Color-naming Stroop task. We used a preprogrammed version of the Stroop task (44) implemented in Apple's ResearchKit task collection and available by default on the MyDataHelps platform. Participants encountered a series of color-related words (red, green, blue, or yellow) displayed in either a congruent or conflicting color. Participants indicated the color of the word, ignoring the text, by selecting the first letter of the color's name, displayed below the word. Participants completed 20 trials daily. We analyzed performance using an individual-level hierarchical DDM of correct/incorrect responses, with days nested within participant (Supplementary Text, note S1), yielding both subject-level parameters for overall drift rate, the effect of congruency on the drift rate, threshold, and nondecision time, as well as daily deviations from these averages for overall drift rate, difficulty effects, and threshold.

Affect-naming Stroop task. We used a custom-written version of the affective Stroop task (45). This task presented participants with a series of basic emotion words (angry, sad, or happy) superimposed on matching or nonmatching angry, sad, or happy face images derived from the Karolinska Directed Emotional Faces database (63). Participants indicated the facial expression displayed in the image, ignoring the text, by selecting the first letter of the emotion's name, displayed below the image. Participants completed 15 trials each day. We analyzed this task in an identical manner to the color-naming Stroop.

Trail-making task. We used a preprogrammed version of the trail-making task (43) available from Apple's ResearchKit task collection. In this task, participants must tap in sequence on a series of circles labeled with numbers (1, 2, 3, ..., 13) or, in a second version, with alternating numbers and letters (1, A, 2, B, 3, C, ..., F, 7). We modeled performance in this task using a hierarchical DDM of taps (coded as correct/incorrect) and response times, similar to the Stroop tasks, with task version (numeric-only or alphanumeric) modifying drift rate as a measure of difficulty.

Go/No-Go task. To assess inhibitory control, we used a variant of the Go/No-Go task with a 4:1 ratio of Go:No-Go cues (orange and blue circles) to maximize false alarms (41). The "Go" action involved tapping a button at the bottom of the screen. Stimuli appeared for 250 ms in random order, with an intertrial interval of 450 ms, during which responses were still recorded. Participants completed 50 trials each day. We used d' (46), corrected for extreme proportions using Hautus' method (64), to assess discrimination accuracy.

Paced serial addition task. We assessed sustained attention and working memory using the Paced Visual Serial Addition Task (42), in which single digits (from 1 to 9) appeared continuously spaced 1 to 3 s apart. Participants had to add each new digit to the one presented immediately before, indicating the correct answer by touching buttons with the digits from 3 to 17 laid out at the bottom of the screen. Participants completed 30 trials daily. Percent accuracy served as our measure of performance.

Spatial span working memory task. As a second measure of working memory and executive function, we used a preprogrammed spatial span working memory task (40) available from Apple's ResearchKit task collection. In this task, participants see a three-by-three grid of flower icons, in which flowers light up sequentially in a pattern that participants must memorize and then repeat. The length of the pattern sequence (the span) is automatically varied during the task, increasing by one after successful sequence recall, and decreasing by one

after a failed attempt. Participants completed a total of six trials, ranging from a sequence length of 2 to 7. Performance on the task is measured by a cumulative score representing the sum of sequence lengths for which the participant provided fully correct recall, multiplied by 20, with a maximum daily task score of 540.

Task motivation and effort questions. In addition to the tasks themselves, we obtained self-reported measures of a priori motivation for completing the tasks and post hoc task load. We measured motivation using a single-item question indicating how motivated the person felt to do well on the tasks, responded to on an 11-point Likert scale running from 0 (not at all motivated) to 10 (extremely motivated) prior to task completion. After task completion, participants completed five items from the NASA-TLX (65), including the degree of mental demand they experienced, temporal demand (e.g., degree of time pressure), how much effort they put into the tasks, how frustrating they found the tasks, and how successful they felt they were at completing the tasks. We also asked two additional questions assessing how distracted the participant felt while completing the tasks, as well as the degree to which they felt their performance was due to luck. Participants responded to these questions on a slider ranging from 0 (low) to 20 (high).

Daily self-reported surveys

To assess aspects of participants' goal-related behaviors, affect, and daily experiences, two sets of surveys were sent out each day. Questions fell into one of four categories: daily feeling states, daily goal pursuit, daily behaviors and experiences, and intended future behavior.

Daily mood/subjective state. We assessed daily mood and other subjective states twice, once in the morning and once in the evening. The morning survey consisted solely of questions about 20 subjective states (8 positive, e.g., calm, happy, excited, motivated, and focused; 12 negative, e.g., depressed, stressed, anxious, angry, sad, lazy, etc.) responded to on a 5-point Likert scale from "not at all" to "extremely." These questions assessed the participant's currently experienced feelings at the start of the day and were available for completion from 5:00 a.m. to 12:30 p.m. The evening survey required participants to complete the same 20 items, only this time indicating how much of each feeling they had experienced "over the past day."

Daily goal pursuit. Our primary measures of daily goal pursuit were derived from the evening survey, which was available for participants to complete from 8:00 p.m. each evening to 4:00 a.m. the next morning. Participants first responded to a general question asking them about their overall I-B gap over the past day on a slider ranging from 0 (no gap) to 100 (accomplishing nothing of what they intended). Participants also responded to a causal attribution question regarding their overall gap, ranging from completely internal (all success or failure due to one's own actions) to completely external (all success or failure due to outside circumstance). They then completed a question regarding how close to ideal their day had been. Participants also completed two questions assessing what percentage (from 0 to 100%) of each of two specific goals they had accomplished. These two goals were derived from questions they had completed the previous day regarding their intended future behavior (see below for details).

Daily behaviors and experiences. Participants completed several questions related to their experiences and behaviors during the day, including how fatigued they had felt, how ill/sick they had felt, and how much time they had spent sleeping, working on primary occupational tasks (e.g., school work), working on nonoccupational tasks (e.g., housework), in solo or nonsolo mental or physical leisure activities, and on social media

(reported in hours) as well as how much time they had spent exercising (reported in minutes). Participants also indicated how much physical effort and mental effort they had expended over the day, how productive they had felt, how much they had procrastinated, how punctual they had been, and whether any unusual events had occurred. Last, participants indicated how many alcoholic drinks they had consumed.

Daily intended future behavior. After reporting on their past day, participants completed a series of questions asking them about their goals for the next day. First, participants described two specific goals. They were encouraged to describe goals that could, in theory, be 100% complete by the following day. Then, for each specific goal, they completed a series of questions about properties of that goal, including how confident they were that they could achieve it, how effortful it would be, how important it was, how extreme the consequences would be if they did not complete it, and how intrinsically and extrinsically motivated they were to complete it. Then, participants answered binary yes/no questions about whether they had goals for the next day regarding how much time they wanted to spend in the following day: primary occupational tasks, nonoccupational tasks, solo leisure, nonsolo leisure, social media, sleep, and exercise. They were also asked whether they held goals regarding drinking behavior. If participants indicated having a goal in a particular domain, they were then prompted to provide a numerical indicator (in hours, minutes, or drinks) for that goal. These questions, combined with questions about daily past behaviors, allowed us to assess the extent of I-B Gaps across a wide variety of dimensions and domains.

Individual difference measures

To determine whether personality influenced associations between daily cognition and behavior, we focus here on results using three of the most popular self-reported measures related to goal achievement and self-reported self-control: the HEXACO Conscientiousness Scale (51), the Brief Self-Control Scale (2), and the Duckworth Grit Scale (52).

Analysis

DDMs of daily performance

We used hierarchical DDM as implemented in the *hssm* package (66) to model daily variation in task performance in three tasks where every response could be categorized as correct/incorrect with a corresponding response time: the color-naming Stroop task, the affective Stroop task, and the trail-making task. For each task and individual, we modeled daily task performance as a function of overall drift rate ν , the effect of trial difficulty (i.e., congruency versus incongruency in the two Stroop tasks and alphanumeric versus numeric-only in the trail-making task) on the drift rate νDfc , the threshold a , and the non-decision time ter . For overall drift ν , νDfc , and a , we fit both day-level and individual average parameters. For ter , we fit a single, overall value for each participant (see the Supplementary Text, note S1 for parameter recovery statistics and model justification).

We then modeled observed data from each of the three tasks, for each individual with at least 21 days of data on a task. To aid in model convergence, we removed the fastest 2% and slowest 1% of each participant's responses. Convergence diagnostics were assessed via the Gelman-Rubin statistic and indicated \hat{r} values <1.01 for between 99.97 and 100% of all individual parameters across tasks.

Removing temporal trends

Consistent with prior work (67), we observed clear effects of learning in many of the cognitive tasks. To remove these effects, we estimated a series of nonlinear hierarchical models predicting a single

task performance parameter y_t (e.g., overall drift rate ν in the color Stroop task, accuracy in the serial addition task, etc.) as a function of the number of times t that the participant had completed the task. To derive an appropriate set of parameters, we visually inspected model residuals, attempting to achieve a reasonable balance between model complexity and the absence of clear temporal trends. This yielded the following parameters for each task performance parameter and subject s :

1) Three parameters ($\alpha_s, \beta_s, \gamma_s$) related to an asymptoting power function of practice, $\alpha_s - \beta_s \times t^{-\gamma_s}$, where α_s represents the participant's asymptotic, steady-state performance level, β_s represents how far below this asymptote the participant begins, and γ_s controls the steepness of the learning curve

2) A linear trend parameter δ_s representing a constant increasing or decreasing effect of repeated practice with the task

We estimated each of these subject-level parameters using Monte Carlo Markov Chain (MCMC) methods as implemented by the *rjags* package in R (68). For each model, we estimated three chains with 1000 adaptation samples, 5000 samples of burn-in, and 10,000 samples from the posterior with a thinning factor of 10. Posterior means for each subject-level parameter were constructed using the *coda* package in R (69).

Detrended values for variation in daily task performance metrics were computed as $r_t = y_t - \hat{y}_t$ where y_t is the observed task value at task repetition time t , and \hat{y}_t represents the predicted subject-level effects of repeated experience. For most tasks, \hat{y}_t was modeled as the posterior mean of a Gaussian distribution $\sim N(\mu_s, \sigma_s)$, where $\mu_{s,t} = \alpha_s - \beta_s \times t^{-\gamma_s} + \delta_s \times t$. In the case of accuracy data, \hat{y}_t was modeled as the output of a β distribution $\sim \beta(a,b)$, where $a = \mu_s \times \phi_s$, and $b = (1 - \mu_s) \times \phi_s$. Inspection of model residuals indicated outliers in the first two to three rounds of task practice, likely due to participants becoming familiar with the task, as well as on the last day or two (days 83 to 84), perhaps due to changes in focus in anticipation of completion. We thus excluded performance on the first two and last 2 days from all analyses (see fig. S3 for two examples of the full preprocessing pipeline), although multiverse analysis suggested that doing so had little effect on conclusions (see Results), and findings remained significant when including them.

Last, we observed temporal trends in self-reported measures of the I-B Gap. To remove these effects, we computed mixed effects linear regressions in which each self-reported gap measure was modeled as a linear function of day in the semester, with subject-specific means and slopes

$$y_{t,s} = \theta_s * t + C_s$$

We then computed residuals of these models in the same manner as the task data, with the detrended results serving as our target measures for analysis.

Linear mixed-effects modeling

To examine correlates of both daily mental sharpness and the I-B Gap, we fit linear mixed-effects models implemented via the *lme4* (70) and *nlme* packages [version 3.1-162; (71, 72)] in R. Unless otherwise noted, we first disaggregated all predictors in these models into between-person intercepts and within-person fluctuations (detrended), using a linear mixed-effects model with day as the predictor. Per the recommendation of Curran and Bauer (50), between-person intercepts were estimated at the middle time point/observation for each individual, and detrended residuals served as the estimate of day-to-day within-person fluctuations. In addition, unless otherwise noted, we normalized both the

dependent and within-person detrended predictors within each participant, resulting in standardized β coefficients for ease of comparison of effects. Because this step removes between-person variance in the outcome measure, we omitted between-person intercepts as predictors in these models. Note, however, that omitting these between-person predictors should not affect estimation of within-person effects, because these predictors are orthogonal by construction. For models using raw (non-normalized) outcome variables (e.g., estimating between-person effects of cognitive performance), we included between-person intercepts estimated from the disaggregation model described above.

For models in which daily mental sharpness served as the dependent variable, residual autocorrelation was explicitly modeled using a third-order autoregressive (AR) structure, which accounted for clear sequential dependencies across consecutive measurement days within each participant. We selected AR(3) based on model comparison using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which indicated superior model fit with this third-order autoregressive structure. Observations containing missing data were excluded via listwise deletion. For models focusing on predictors of the I-B gap, we used an AR(1) autocorrelation structure based on similar model-comparison metrics.

Multiverse analysis

To determine the robustness of our findings, we engaged in a multiverse analysis (53), analyzing the relationship between the I-B gap and daily mental sharpness under different assumptions and approaches. We explored every combination of the following alternatives: (i) We examined data from only the first 3 weeks of the study versus all 84 days; (ii) we examined data when including only participants meeting one of four different cutoffs for total number of observations (no minimum, or a minimum of 16%, 33%, or 50% valid observations); (iii) we examined data with and without trimming the first and last days, including trimming of 0, 2, or 7 days from the beginning and end of the study; and (iv) we examined six different variants of the daily mental sharpness composite measure. In addition to the version reported in the primary analyses (referred to here and in Fig. 5 as CC1), we tested the following alternatives: (i) a version in which only the tasks primarily tapping working memory (PSAT and spatial working memory) were included (CC2); (ii) a version in which only parameters from classic conflict tasks (i.e., Go/No-Go d' , color Stroop overall drift, and affective Stroop overall drift) were included (CC3); (iii) a version in which all six tasks were included, but overall accuracy measures (rather than drift rates) were used as cognitive parameters from the color Stroop, affective Stroop, and trail-making tasks (CC4); (iv) a version including only traditional metrics from the three conflict tasks (CC5: Go/No-Go d' , overall accuracy in the color Stroop, and overall accuracy in the affective Stroop); and, last, (v) a version using an average of the same six task parameters as in CC1, but without any removal of practice effects (CC6). This resulted in 144 unique combinations of methodological and analytical decisions. For each of these 144 combinations, we then conducted a mixed-effects AR(1) model predicting the I-B gap from the appropriate CC (as defined above), with the specified exclusions and data preprocessing.

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S3
References

REFERENCES

1. T. E. Moffitt, L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, A. Caspi, A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2693–2698 (2011).
2. J. P. Tangney, R. F. Baumeister, A. L. Boone, High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *J. Pers.* **72**, 271–324 (2004).
3. A. L. Duckworth, D. Weir, E. Tsukayama, D. Kwok, Who does well in life? Conscientious adults excel in both objective and subjective success. *Front. Psychol.* **3**, 356 (2012).
4. M. P. Wilmot, D. S. Ones, A century of research on conscientiousness at work. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23004–23010 (2019).
5. I. W. Eisenberg, P. G. Bissett, A. Zeynep Enkavi, J. Li, D. P. MacKinnon, L. A. Marsch, R. A. Poldrack, Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* **10**, 2319 (2019).
6. M. Inzlicht, B. W. Roberts, The fable of state self-control. *Curr. Opin. Psychol.* **58**, 101848 (2024).
7. A. L. Duckworth, The significance of self-control. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2639–2640 (2011).
8. D. T. D. de Ridder, G. Lensvelt-Mulders, C. Finkenauer, F. M. Stok, R. F. Baumeister, Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Pers. Soc. Psychol. Rev.* **16**, 76–99 (2012).
9. R. F. Baumeister, K. D. Vohs, D. M. Tice, The strength model of self-control. *Curr. Dir. Psychol. Sci.* **16**, 351–355 (2007).
10. E. C. Carter, L. M. Kofler, D. E. Forster, M. E. McCullough, A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *J. Exp. Psychol. Gen.* **144**, 796–815 (2015).
11. M. Friese, D. D. Loschelder, K. Gieseler, J. Frankenbach, M. Inzlicht, Is ego depletion real? An analysis of arguments. *Pers. Soc. Psychol. Rev.* **23**, 107–131 (2019).
12. M. S. Hagger, N. L. D. Chatzisarantis, H. Alberts, C. O. Anggono, C. Batailler, A. R. Birt, R. Brand, M. J. Brandt, G. Brewer, S. Bruyneel, D. P. Calvillo, W. K. Campbell, P. R. Cannon, M. Carlucci, N. P. Carruth, T. Cheung, A. Crowell, D. T. D. De Ridder, S. Dewitte, M. Elson, J. R. Evans, B. A. Fay, B. M. Fennis, A. Finley, Z. Francis, E. Heise, H. Hoemann, M. Inzlicht, S. L. Koole, L. Koppel, F. Kroese, F. Lange, K. Lau, B. P. Lynch, C. Martijn, H. Merckelbach, N. V. Mills, A. Michirev, A. Miyake, A. E. Mosser, M. Muise, D. Muller, M. Muzi, D. Nalis, N. Nurwanti, H. Otgaar, M. C. Philipp, P. Primoceri, K. Rentzsch, L. Ringos, C. Schlinkert, B. J. Schmeichel, S. F. Schoch, M. Schrama, A. Schütz, A. Stamos, G. Tinghög, J. Ullrich, M. vanDellen, S. Wimbari, W. Wolff, C. Yusainy, O. Zerhouni, M. Zwienerberg, A multilab preregistered replication of the ego-depletion effect. *Pers. Psychol. Sci.* **11**, 546–573 (2016).
13. M. Inzlicht, B. J. Schmeichel, C. N. Macrae, Why self-control seems (but may not be) limited. *Trends Cogn. Sci.* **18**, 127–133 (2014).
14. A. Weigard, D. A. Clark, C. Sripatha, Cognitive efficiency beats top-down control as a reliable individual difference dimension relevant to self-control. *Cognition* **215**, 104818 (2021).
15. C. Löffler, G. T. Frischkorn, D. Hagemann, K. Sadus, A.-L. Schubert, The common factor of executive functions measures nothing but speed of information uptake. *Psychol. Res.* **88**, 1092–1114 (2024).
16. E. T. Berkman, E. B. Falk, M. D. Lieberman, In the trenches of real-world self-control: Neural correlates of breaking the link between craving and smoking. *Psychol. Sci.* **22**, 498–506 (2011).
17. L. Wennerhold, M. Friese, Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra Psychol.* **6**, 9 (2020).
18. W. Hofmann, M. Adriaanse, K. D. Vohs, R. F. Baumeister, Dieting and the self-control of eating in everyday environments: An experience sampling study. *Br. J. Health Psychol.* **19**, 523–539 (2014).
19. K.-M. Krönke, M. Wolff, H. Mohr, A. Kräplin, M. N. Smolka, G. Bühringer, T. Goschke, Monitor yourself! Deficient error-related brain activity predicts real-life self-control failures. *Cogn. Affect. Behav. Neurosci.* **18**, 622–637 (2018).
20. G. D. Logan, R. J. Schachar, R. Tannock, Impulsivity and inhibitory control. *Psychol. Sci.* **8**, 60–64 (1997).
21. P. G. Enticott, J. R. P. O'gloff, J. L. Bradshaw, Associations between laboratory measures of executive inhibitory control and self-reported impulsivity. *Pers. Individ. Dif.* **41**, 285–294 (2006).
22. B. Saunders, M. Milyavskaya, A. Etz, D. Randles, M. Inzlicht, Reported self-control is not meaningfully associated with inhibition-related executive function: A Bayesian analysis. *Collabra Psychol.* **4**, 39 (2018).
23. B. Saunders, M. Milyavskaya, M. Inzlicht, Longitudinal evidence that event related potential measures of self-regulation do not predict everyday goal pursuit. *Nat. Commun.* **13**, 3201 (2022).
24. C. D. Von Gunten, B. D. Bartholow, J. S. Martins, Inhibition tasks are not associated with a variety of behaviors in college students. *Eur. J. Pers.* **34**, 412–430 (2020).
25. S. V. Burks, J. P. Carpenter, L. Goette, A. Rustichini, Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7745–7750 (2009).

26. P. C. M. Molenaar, A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Meas. Interdiscip. Res. Perspect.* **2**, 201–218 (2004).
27. D. Borsboom, G. J. Mellenbergh, J. van Heerden, The theoretical status of latent variables. *Psychol. Rev.* **110**, 203–219 (2003).
28. A. J. Fisher, J. D. Medaglia, B. F. Jeronimus, Lack of group-to-individual generalizability is a threat to human subjects research. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6106–E6115 (2018).
29. R. S. Stawski, M. J. Sliwinski, S. M. Hofer, Between-person and within-person associations among processing speed, attention switching, and working memory in younger and older adults. *Exp. Aging Res.* **39**, 194–214 (2013).
30. K. D. Vohs, B. J. Schmeichel, S. Lohmann, Q. F. Gronau, A. J. Finley, S. E. Ainsworth, J. L. Alquist, M. D. Baker, A. Brizi, A. Bunyi, G. J. Butschek, C. Campbell, J. Capaldi, C. Cau, H. Chambers, N. L. D. Chatzisarantis, W. J. Christensen, S. L. Clay, J. Curtis, V. De Cristofaro, K. del Rosario, K. Diel, Y. Doğruoğlu, M. Doi, T. L. Donaldson, A. B. Eder, M. Ersoff, J. R. Eyink, A. Falkenstein, B. M. Fennis, M. B. Findley, E. J. Finkel, V. Forgea, M. Frieze, P. Fuglestad, N. E. Garcia-Willingham, L. F. Geraedts, W. M. Gervais, M. Giacomantonio, B. Gibson, K. Gieseler, J. Gineikiene, E. M. Gloger, C. M. Gobes, M. Grande, M. S. Hagger, B. Hartsell, A. D. Hermann, J. J. Hidding, E. R. Hirt, J. Hodge, W. Hofmann, J. L. Howell, R. D. Hutton, M. Inzlicht, L. James, E. Johnson, H. L. Johnson, S. M. Joyce, Y. Joye, J. H. Kaben, L. K. Kammrath, C. N. Kelly, B. L. Kissell, S. L. Koole, A. Krishna, C. Lam, K. T. Lee, N. Lee, D. C. Leighton, D. D. Loschelder, H. M. Maranges, E. J. Masicampo, K. Mazara, S. McCarthy, I. McGregor, N. L. Mead, W. B. Mendes, C. Meslot, N. M. Michalak, M. Milyavskaya, A. Miyake, M. Moeini-Jazani, M. Muraven, E. Nakahara, K. Patel, J. V. Petrocelli, K. M. Pollak, M. M. Price, H. J. Ramsey, M. Rath, J. A. Robertson, R. Rockwell, I. F. Russ, M. Salvati, B. Saunders, A. Scherer, A. Schütz, K. N. Schmitt, S. C. Segerstrom, B. Serenka, K. Sharpinsky, M. Shaw, J. Sherman, Y. Song, N. Sosa, K. Spillane, J. Stapels, A. R. Stinnett, H. R. Strawser, K. Sweeny, D. Theodore, K. Tonn, Y. van Oldenbeuving, M. R. van Dellen, R. C. Vergara, J. S. Walker, C. E. Waugh, F. Weise, K. M. Werner, C. Wheeler, R. A. White, A. L. Wichman, B. J. Wiggins, J. A. Wills, J. H. Wilson, E.-J. Wagenmakers, D. Albarracín, A multisite preregistered paradigmatic test of the ego-depletion effect. *Psychol. Sci.* **32**, 1566–1581 (2021).
31. K. E. Garrison, A. J. Finley, B. J. Schmeichel, Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments. *Pers. Soc. Psychol. Bull.* **45**, 728–739 (2019).
32. B. Blain, G. Hollard, M. Pessiglione, Neural mechanisms underlying the impact of daylong cognitive work on economic decisions. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6967–6972 (2016).
33. F. Blume, A. Irmer, J. Dirk, F. Schmiedek, Day-to-day variation in students' academic success: The role of self-regulation, working memory, and achievement goals. *Dev. Sci.* **25**, e13301 (2022).
34. D. J. H. Powell, D. McMinn, J. L. Allan, Does real time variability in inhibitory control drive snacking behavior? An intensive longitudinal study. *Health Psychol.* **36**, 356–364 (2017).
35. K. E. Smith, T. B. Mason, L. M. Schaefer, A. Juarascio, R. Dvorak, N. Weinbach, R. D. Crosby, S. A. Wonderlich, Examining intra-individual variability in food-related inhibitory control and negative affect as predictors of binge eating using ecological momentary assessment. *J. Psychiatr. Res.* **120**, 137–143 (2020).
36. A. Jones, B. Tiplady, K. Houben, C. Nederkoorn, M. Field, Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. *Psychopharmacology (Berl)* **235**, 1487–1496 (2018).
37. W. Wolff, C. S. Martarelli, Bored into depletion? Toward a tentative integration of perceived self-control exertion and boredom as guiding signals for goal-directed behavior. *Pers. Psychol. Sci.* **15**, 1272–1283 (2020).
38. R. Kurzban, A. Duckworth, J. W. Kable, J. Myers, An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* **36**, 661–679 (2013).
39. S. Lyubomirsky, K. M. Sheldon, D. Schkade, Pursuing happiness: The architecture of sustainable change. *Rev. Gen. Psychol.* **9**, 111–131 (2005).
40. R. P. C. Kessels, M. J. E. van Zandvoort, A. Postma, L. J. Kappelle, E. H. F. de Haan, The Corsi Block-Tapping Task: Standardization and normative data. *Appl. Neuropsychol.* **7**, 252–258 (2000).
41. M. E. Young, S. C. Sutherland, A. W. McCoy, Optimal go/no-go ratios to maximize false alarms. *Behav. Res. Methods* **50**, 1020–1029 (2018).
42. T. N. Tombaugh, A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Arch. Clin. Neuropsychol.* **21**, 53–76 (2006).
43. K. Arbutnot, J. Frank, Trail Making Test, Part B as a measure of executive control: Validation using a set-switching paradigm. *J. Clin. Exp. Neuropsychol.* **22**, 518–528 (2000).
44. J. R. Stroop, Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662 (1935).
45. A. Etkin, T. Egner, D. M. Peraza, E. R. Kandel, J. Hirsch, Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* **51**, 871–882 (2006).
46. H. Stanislaw, N. Todorov, Calculation of signal detection theory measures. *Behav. Res. Methods, Instr. Comp.* **31**, 137–149 (1999).
47. T. V. Wiecki, I. Sofer, M. J. Frank, HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Front. Neuroinform.* **7**, 14 (2013).
48. F. M. Lord, M. R. Novick, *Statistical Theories of Mental Test Scores* (Information Age Publishing, Inc., 2008).
49. A. Brose, F. Schmiedek, M. Lövdén, U. Lindenberger, Daily variability in working memory is coupled with negative affect: the role of attention and motivation. *Emotion*, **12**, 605 (2012).
50. P. J. Curran, D. J. Bauer, The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.* **62**, 583–619 (2011).
51. M. Ashton, K. Lee, The HEXACO-60: A short measure of the major dimensions of personality. *J. Pers. Assess.* **91**, 340–345 (2009).
52. A. L. Duckworth, C. Peterson, M. D. Matthews, D. R. Kelly, Grit: Perseverance and passion for long-term goals. *J. Pers. Soc. Psychol.* **92**, 1087–1101 (2007).
53. S. Steegen, F. Tuerlinckx, A. Gelman, W. Vanpaemel, Increasing transparency through a multiverse analysis. *Respect. Psychol. Sci.* **11**, 702–712 (2016).
54. K. M. Werner, M. Inzlicht, B. Q. Ford, Whither inhibition? *Curr. Dir. Psychol. Sci.* **31**, 333–339 (2022).
55. E. T. Berkman, C. A. Hutcherson, J. L. Livingston, L. E. Kahn, M. Inzlicht, Self-control as value-based choice. *Curr. Dir. Psychol. Sci.* **26**, 422–428 (2017).
56. W. Hofmann, B. J. Schmeichel, A. D. Baddeley, Executive functions and self-regulation. *Trends Cogn. Sci.* **16**, 174–180 (2012).
57. R. B. Lopez, W. Hofmann, D. D. Wagner, W. M. Kelley, T. F. Heatherton, Neural predictors of giving in to temptation in daily life. *Psychol. Sci.* **25**, 1337–1344 (2014).
58. H. J. Lee, D. M. Smith, C. E. Hauenstein, A. Dworetzky, B. T. Kraus, M. Dorn, D. E. Nee, C. Gratton, Precise individual measures of inhibitory control. *Nat. Hum. Behav.* **9**, 1613–1630 (2025).
59. T. Müller, M. C. Klein-Flügge, S. G. Manohar, M. Husain, M. A. J. Apps, Neural and computational mechanisms of momentary fatigue and persistence in effort-based choice. *Nat. Commun.* **12**, 4593 (2021).
60. P. Klasnja, E. B. Hekler, S. Shiffman, A. Boruvka, D. Almirall, A. Tewari, S. A. Murphy, Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychol.* **34**, 1220–1228 (2015).
61. A. Smyth, M. Milyavskaya, M. Frieze, K. Werner, M.-L. Frech, D. Loschelder, J. Anderson, M. Inzlicht, M. Kolbuszewska, K. Wang, What constitutes successful goal pursuit? Exploring the relation between subjective and objective measures of goal progress. *Pers. Sci.* **4**, 10.5964/ps.12017 (2023).
62. S. Stier, J. Breuer, P. Siegers, K. Thorson, Integrating survey data and digital trace data: Key issues in developing an emerging field. *Soc. Sci. Comput. Rev.* **38**, 503–516 (2020).
63. D. Lundqvist, A. Flykt, A. Öhman, The Karolinska Directed Emotional Faces (KDEF), APA PsycTests (1998); <https://doi.org/10.1037/t27732-000>.
64. M. J. Hautus, Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behav. Res. Methods Instr. Comp.* **27**, 46–51 (1995).
65. S. G. Hart, L. E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* **52**, 139–183 (1988).
66. A. Fengler, Y. Xu, K. Bera, A. Omar, M. J. Frank, HSSM: A generalized toolbox for hierarchical Bayesian estimation of computational models in cognitive neuroscience, (2025); <https://Inccbrown.github.io/HSSM/>.
67. R. Schurr, D. Reznik, H. Hillman, R. Bhui, S. J. Gershman, Dynamic computational phenotyping of human cognition. *Nat. Hum. Behav.* **8**, 917–931 (2024).
68. M. Plummer, rjags: Bayesian graphical models using MCMC. R package version 4-15 (2023); <https://CRAN.R-project.org/package=rjags>.
69. M. Plummer, N. Best, K. Cowles, K. Vines, CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).
70. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
71. J. C. Pinheiro, D. M. Bates, "Linear mixed-effects models: Basic concepts and examples" in *Mixed-Effects Models in S and S-PLUS. Statistics and Computing* (Springer-Verlag, 2000), pp. 3–56.
72. J. C. Pinheiro, D. Bates, nlme: Linear and nonlinear mixed effects models. R package version 3.1-164 (2023).
73. J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
74. G. M. Becker, M. H. Degroot, J. Marschak, Measuring utility by a single-response sequential method. *Behav. Sci.* **9**, 226–232 (1964).
75. C. W. Lejuez, J. P. Read, C. W. Kahler, J. B. Richards, S. E. Ramsey, G. L. Stuart, D. R. Strong, R. A. Brown, Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* **8**, 75–84 (2002).
76. R. C. Moore, M. A. Hussain, C. W.-M. Watson, P. L. Fazeli, M. J. Marquine, B. C. Yarns, D. V. Jeste, D. J. Moore, Grit and ambition are associated with better neurocognitive and everyday functioning among adults living with HIV. *AIDS Behav.* **22**, 3214–3225 (2018).
77. A. A. Struk, J. S. A. Carriere, J. A. Cheyne, J. Danckert, A short boredom proneness scale. *Assessment* **24**, 346–359 (2017).
78. R. H. Hoyle, M. T. Stephenson, P. Palmgreen, E. P. Lorch, R. L. Donohew, Reliability and validity of a brief measure of sensation seeking. *Pers. Individ. Dif.* **32**, 401–414 (2002).

79. E. L. Deci, R. M. Ryan, The general causality orientations scale: Self-determination in personality. *J. Res. Pers.* **19**, 109–134 (1985).
80. J. Houston, P. Harris, S. McIntire, D. Francis, Revising the competitiveness index using factor analysis. *Psychol. Rep.* **90**, 31–34 (2002).
81. P. F. Lovibond, S. H. Lovibond, The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Therapy* **33**, 335–343 (1995).
82. E. Diener, D. Wirtz, W. Tov, C. Kim-Prieto, D. Choi, S. Oishi, R. Biswas-Diener, New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Soc. Indic. Res.* **97**, 143–156 (2010).
83. L. L. Carstensen, F. R. Lang, Future Time Perspective Scale (2018); <https://doi.org/10.1037/t31314-000>.
84. R. Schwarzer, M. Jerusalem, “Generalized self-efficacy scale” in *Measures in Health Psychology: A User’s Portfolio. Causal and Control Beliefs*, J. Weinman, S. Wright, M. Johnston, Eds. (NFER-NELSON, 1995), pp. 35–37.
85. O. N. E. Kjell, D. Daukantaitė, K. Hefferon, S. Sikström, The Harmony in Life Scale complements the Satisfaction with Life Scale: Expanding the conceptualization of the cognitive component of subjective well-being. *Soc. Indic. Res.* **126**, 893–919 (2016).
86. H. Cantril, *The Pattern of Human Concerns* (Rutgers University Press, 1965).
87. K. W. Brown, R. M. Ryan, The benefits of being present: Mindfulness and its role in psychological well-being. *J. Pers. Soc. Psychol.* **84**, 822–848 (2003).
88. J. T. Cacioppo, R. E. Petty, The need for cognition. *J. Pers. Soc. Psychol.* **42**, 116–131 (1982).
89. S. Cohen, T. Kamarck, R. Mermelstein, A global measure of perceived stress. *J. Health Soc. Behav.* **24**, 385–396 (1983).
90. D. Watson, L. A. Clark, A. Tellegen, Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988).
91. R. C. Kessler, P. R. Barker, L. J. Colpe, J. F. Epstein, J. C. Gfroerer, E. Hiripi, M. J. Howes, S.-L. T. Normand, R. W. Manderscheid, E. E. Walters, A. M. Zaslavsky, Screening for serious mental illness in the general population. *Arch. Gen. Psychiatry* **60**, 184–189 (2003).
92. K. E. Stanovich, R. F. West, On the relative independence of thinking biases and cognitive ability. *J. Pers. Soc. Psychol.* **94**, 672–695 (2008).
93. J. G. Lynch, R. G. Netemeyer, S. A. Spiller, A. Zammit, A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *J. Cons. Res.* **37**, 108–128 (2010).
94. C. S. Burckhardt, K. L. Anderson, The Quality of Life Scale (QOLS): Reliability, validity, and utilization. *Health Qual. Life Outcomes* **1**, 60 (2003).
95. E. Diener, R. A. Emmons, R. J. Larsen, S. Griffin, The Satisfaction with Life Scale. *J. Pers. Assess.* **49**, 71–75 (1985).
96. G. Block, C. Gillespie, E. H. Rosenbaum, C. Jensen, A rapid food screener to assess fat and fruit and vegetable intake. *Am. J. Prev. Med.* **18**, 284–288 (2000).
97. M. Rosenberg, *Society and the Adolescent Self-Image* (Princeton University Press, 1965); 10.1515/9781400876136.
98. R. W. Robins, H. M. Hendin, K. H. Trzesniewski, Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Pers. Soc. Psychol. Bull.* **27**, 151–161 (2001).
99. N. Xanidis, C. M. Brignell, The association between the use of social network sites, sleep quality and cognitive function during the day. *Comput. Human Behav.* **55**, 121–126 (2016).
100. M. Kiernan, D. E. Schoffman, K. Lee, S. D. Brown, J. M. Fair, M. G. Perri, W. L. Haskell, The Stanford Leisure-Time Activity Categorical Item (L-Cat): A single categorical item sensitive to physical activity changes in overweight/obese women. *Int. J. Obes.* **37**, 1597–1602 (2013).
101. S. Lyubomirsky, H. S. Lepper, A measure of subjective happiness: Preliminary reliability and construct validation. *Soc. Indic. Res.* **46**, 137–155 (1999).
102. W. M. Reynolds, Development of reliable and valid short forms of the marlowe-crowne social desirability scale. *J. Clin. Psychol.* **38**, 119–125 (1982).
103. P. Salovey, J. D. Mayer, Emotional intelligence. *Imagin. Cogn. Pers.* **9**, 185–211 (1990).
104. K. Bernecker, D. Becker, Beyond self-control: Mechanisms of hedonic goal pursuit and its relevance for well-being. *Pers. Soc. Psychol. Bull.* **47**, 627–642 (2021).
105. S. Beattie, L. Hardy, J. Savage, T. Woodman, N. Callow, Development and validation of a trait measure of robustness of self-confidence. *Psychol. Sport Exer.* **12**, 184–191 (2011).
106. G. Blau, J. Ryan, On measuring work ethic: A neglected work commitment facet. *J. Vocat. Behav.* **51**, 435–448 (1997).
107. J. Fan, B. McCandliss, J. Fossella, J. Flombaum, M. Posner, The activation of attentional networks. *Neuroimage* **26**, 471–479 (2005).
108. E. T. Cokely, M. Galesic, E. Schulz, S. Ghazal, R. Garcia-Retamero, Measuring risk literacy: The Berlin Numeracy Test. *Judgm. Decis. Mak.* **7**, 25–47 (2012).
109. M. N. Koffarnus, W. K. Bickel, A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Exp. Clin. Psychopharmacol.* **22**, 222–228 (2014).
110. T. Shallice, M. E. Evans, The involvement of the frontal lobes in cognitive estimation. *Cortex* **14**, 294–303 (1978).
111. A. Westbrook, D. Kester, T. S. Braver, What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLOS ONE* **8**, e68210 (2013).
112. M. E. Toplak, R. F. West, K. E. Stanovich, Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Think. Reason.* **20**, 147–168 (2014).
113. K. N. Ochsner, R. D. Ray, J. C. Cooper, E. R. Robertson, S. Chopra, J. D. E. Gabrieli, J. J. Gross, For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage* **23**, 483–499 (2004).
114. C. A. Holt, S. K. Laury, Risk aversion and incentive effects. *Am. Econ. Rev.* **92**, 1644–1655 (2002).
115. L. Clark, T. W. Robbins, K. D. Ersche, B. J. Sahakian, Reflection impulsivity in current and former substance users. *Biol. Psychiatry* **60**, 515–522 (2006).
116. M. A. Lawson, R. P. Larrick, J. B. Soll, Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules. *Judgm. Decis. Mak.* **15**, 660–684 (2020).
117. D. B. Yntema, Keeping track of several things at once. *Hum. Factors* **5**, 7–17 (1963).
118. R. Hull, R. C. Martin, M. E. Beier, D. Lane, A. C. Hamilton, Executive function in older adults: A structural equation modeling approach. *Neuropsychology* **22**, 508–522 (2008).
119. J. Raven, J. Raven, “Raven Progressive Matrices” in *Handbook of Nonverbal Assessment* (Springer, 2003), pp. 223–237.
120. G. D. Roach, D. Dawson, N. Lamond, Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiol. Int.* **23**, 1379–1387 (2006).
121. Z. Shao, E. Janse, K. Visser, A. S. Meyer, What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Front. Psychol.* **5**, 772 (2014).
122. R. C. Wilson, Y. Niv, Inferring relevance in a changing world. *Front. Hum. Neurosci.* **5**, 189 (2012).
123. C. Lu, R. W. Proctor, The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychon. Bull. Rev.* **2**, 174–207 (1995).
124. R. Krikorian, J. Bartok, N. Gay, Tower of London procedure: A standard method and developmental data. *J. Clin. Exp. Neuropsychol.* **16**, 840–850 (1994).
125. J. Gläscher, N. Daw, P. Dayan, J. P. O’Doherty, States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).

Acknowledgments: We thank A. Jeyapalan and Y. Wei for assistance in recruitment and onboarding; J. Oh, K. Lubana, and M. Murugesan for assistance in coding microtasks and creating an experiment completion dashboard in MyDataHelps; F. Faisal, H. Reeyaz, J. Zhang, and J. Guan for assistance with categorization and coding of goals; and M. Inzlicht and members of the Decision Neuroscience Laboratory for helpful comments on earlier versions of this manuscript. **Funding:** This work was supported by a Canada Research Chair Award from the Natural Sciences and Engineering Research Council to C.A.H.; a Discovery Grant from the Natural Sciences and Engineering Research Council (RGPIN-2016-05641) to C.A.H.; and a Natural Sciences and Engineering Research Council Post-Graduate Scholarship to D.J.W. **Author contributions:** Conceptualization: D.J.W. and C.A.H. Methodology: D.J.W. and C.A.H. Software: D.J.W. and C.A.H. Validation: D.J.W. and C.A.H. Formal analysis: D.J.W. and C.A.H. Investigation: D.J.W. and C.A.H. Resources: C.A.H. Data curation: D.J.W. and C.A.H. Writing—original draft: C.A.H. Writing—review and editing: D.J.W. and C.A.H. Visualization: C.A.H. Supervision: C.A.H. Project administration: D.J.W. and C.A.H. Funding acquisition: C.A.H. **Competing interests:** The authors declare that they have no competing interest. **Data and materials availability:** Data and code necessary to reproduce the results reported in this paper are available at <https://doi.org/10.17605/OSF.IO/9X64V>. All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. This study did not generate new materials.

Submitted 23 July 2025
Accepted 5 January 2026
Published 4 February 2026
10.1126/sciadv.aea8697

Day-to-day fluctuations in cognitive precision predict the domain-general intention-behavior gap

Daniel J. Wilson and Cendri A. Hutcherson

Sci. Adv. **12** (6), eaea8697. DOI: 10.1126/sciadv.aea8697

View the article online

<https://www.science.org/doi/10.1126/sciadv.aea8697>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).