Journal of Experimental Psychology: General

A Watched Pot Seems Slow to Boil: Why Frequent Monitoring Decreases Perceptions of Progress

André Vaz, André Mata, and Clayton R. Critcher

Online First Publication, February 17, 2025. https://dx.doi.org/10.1037/xge0001733

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001733

A Watched Pot Seems Slow to Boil: Why Frequent Monitoring Decreases Perceptions of Progress

André Vaz[1, 2], André Mata[2], and Clayton R. Critcher[3]
[1] Department of Psychology, Ruhr University Bochum
[2] CICPSI, Faculdade de Psicologia, Universidade de Lisboa
[3] Haas School of Business, University of California, Berkeley

In evaluating changing attributes (e.g., work output, pollution levels), perceivers care not only about an attribute's level but its rate of change. Two employees likely have different value in the eyes of a supervisor if they take different amounts of time to complete the same work. Ten studies in the main article (and five in the Supplemental Materials) document and explore a monitoring frequency effect (MFE): Progress is seen to slow to the extent it is monitored more frequently. This effect was observed across various domains (workplace, public health, environmental, investment, physical growth) and was robust to financial incentives that encouraged accuracy. Several factors are identified that affect preferences for monitoring targets more or less frequently. Participants also displayed preferences for how frequently they themselves would be monitored; this investigation directly revealed the counterintuitive nature of the MFE. Although the MFE was robust to all tested variants, the size of the MFE did depend on how information about attribute changes was presented. Two mechanistic accounts—one rooted in memory biases for tracked information and the other in a failure to synthesize the tracked information in a normative way—were tested. Only the latter was supported. Discussion focuses on how the MFE complements or only superficially contradicts previous work on myopic loss aversion, the ratio bias, partition dependence, and tracking goal progress. The MFE identifies a qualitatively distinct way by which prior evaluations and beliefs can color evaluations of targets, thereby reinforcing even misguided priors.

Public Significance Statement

In many contexts, perceivers track how quickly change is occurring. Supervisors track employee output, investors track their portfolios' growth, anxious members of the public track the spread of a novel contagious disease. Those perceivers then make decisions based on perceived rates of change. Supervisors promote productive employees, investors reallocate assets toward promising opportunities, and members of the public take steps to address perceived health threats. Although perceptions of attributes' rates of change are (and should be) a function of those attributes' actual rates of change, the present research shows they are also distorted by an action of the perceiver: how often they are engaging in monitoring. The present work identifies several ways these judgment distortions can be reduced.

Keywords: attribute change, temporal neglect, evaluation by moments, productivity

Supplemental materials: https://doi.org/10.1037/xge0001733.supp

Abigail Sussman served as action editor.
André Vaz https://orcid.org/0000-0003-3352-5455
Materials, data, and analysis code are available on the Open Science Framework at https://osf.io/kt8rj/?view_only=3722dd9d1edd442aab4fc2b24a20de27. Clayton R. Critcher received funding from Award 1749608 from the U.S. National Science Foundation. This research was funded in whole, or in part, by the Foundation for Science and Technology of Portugal (Grants PD/BD/150547/2019 and PTDC/PSI-GER/7592/2020). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

André Vaz played an equal role in conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing. André Mata played an equal role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, supervision, visualization, writing–original draft, and writing–review and editing. Clayton R. Critcher played an equal role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, supervision, visualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to André Vaz, Department of Psychology, Ruhr University Bochum, Universitätsstraße 150, 44801, Bochum, Germany. Email: Andre.Vaz@ruhr-uni-bochum.de or andre.vaz@rub.de

To keep tabs on one's world, it is necessary to track rates of change. Employees get work done, but just how quickly? Countries release pollutants into the atmosphere, but at how alarming of a rate? In monitoring change, people typically have a choice over how often they check in to learn whether and how quickly change is occurring. A conscientious home financial manager will consult their credit card statement to see how quickly their balance is building. An overbearing factory supervisor may make the rounds more often to inspect the latest output. Anxious academics may keep a close eye on their cohort's Google Scholar citations, being attuned to the latest upticks in their metrics.

On the one hand, keeping a watchful eye on change may seem tantamount to being fully in the know. The surest way of not knowing whether one has gained weight over the holidays is to stay off the scale. But in this article, we instead consider some unforeseen consequences of the frequency with which one tracks progress. Drawing on social psychology, cognitive psychology, and judgment and decision-making literatures, we hypothesize that frequent (vs. infrequent) monitoring encourages the perception that progress is happening more slowly. In the process, we show not only that the watched pot *seems* slow to boil, but also explore why that is the case, identify several determinants of what leads (proverbial) cooks to keep a more watchful eye, and assess whether people readily intuit the broad applicability of this proverb or instead display counterproductive preferences for how frequently they would want to be monitored.

A brief example will both capture the intuition and provide an introduction to our basic paradigm. Consider a factory supervisor whose job is to monitor factoryworkers' output. Every so often, the supervisor stops by a worker's station to surreptitiously check the worker's output, counting how many parts have been completed and placed in an output tray before clearing the tray's contents until the next time the supervisor completes an inspection. Even if two workers produce output at the same rate, what the supervisor will see in the tray (a straightforward and salient index of output) will be larger or smaller depending on how frequently they check in. If the supervisor, perhaps more concerned about one of the workers than the other, finds themselves stopping by one worker's station twice as often as the other's, then they will always see half of the output in this frequently monitored worker's tray than they would have seen under the less frequent monitoring schedule. To the extent that the supervisor (at least partially) neglects how much time has passed between check-ins, then they may come to see the more frequently monitored worker as less productive. This is one instantiation of what we call the monitoring frequency effect (MFE).

More generally, we suggest that progress is often monitored by considering how much has changed since a previous check-in. Dieters who do weekly weigh-ins may focus more on how much weight they have lost rather than their absolute weights. Public health departments may release the number of new COVID-19 infections instead of simply updating a running tally of total infections. Investors may take note of percentage changes in the market instead of their preferred stock index's current value. PhD advisors may focus more on how many more pages of a draft their advisee sent over instead of on the current length of that article. Given monitoring progress entails attention to, well, progress—or incremental change—the possibility that people may not draw normative conclusions based on how much progress they observe has potentially broad implications.

To understand how quickly progress is happening, two variables are relevant. One is an observed change in an attribute value ("I am down 2 pounds …"). The other is the time course over which that change was observed ("… since I weighed myself a week ago"). Hsee et al. (2019) argued that many psychological biases stem from *relevance insensitivity*, disproportionate reliance on a diagnostic target variable that fails to adjust for variation in situational variables that shape how that diagnostic cue should be interpreted (Yang et al., 2021). Through this lens, observed change is the focal target variable, because it is the direct index of progress or change. In contrast, the passage of time is the situational variable that *should* change the meaning of the observed change (into a rate of change) but is likely to be relatively neglected. After all, the passage of time is just that—an always-present background feature that occurs regardless of whether there is change to be tracked or not. The fact that in numerous contexts change is tracked in standardized intervals (e.g., public companies' quarterly reports, weight loss programs' weekly meetings, governmental agencies' annual updates) only reinforces that tracking the rate of progress demands more sensitivity to observed changes than to the time intervals over which those changes were recorded. If there is no variability in monitoring frequency, only the variability in observed changes is relevant. The passage of time is thus not only an easily neglected situational feature, but one that sometimes *can* be ignored.

Although we are aware of no previous exploration of the MFE, there is precedent for the idea that people neglect scope or duration when making judgments and evaluations. More generally, people focus on salient foreground information that can be gleaned from a singular representation or at a single moment in time and neglect background information that becomes relevant across contexts and periods (Li & Hsee, 2021; Read et al., 1999; Tversky & Kahneman, 1992). When considering how much to donate to a cause, people focus on their concern for the problem and neglect just how extensive (e.g., how many people are affected) the target of assistance is (Hsee et al., 2013). When considering a risky gamble, people focus on what can be gained or lost in the moment and fail to consider the consequences in terms of overall wealth (Kahneman & Tversky, 1979). When considering an annual expense, people find it subjectively less costly to think how much they would be spending per day, because they fail to internalize just how much such small payments add up over time (Atlas & Bartels, 2018; Gourville, 1998). In terms of the MFE, people may be responsive to focal updates of how much change has occurred and fail to appreciate that a full understanding of the rate of change requires one to appreciate the temporal period or scope to which that update applies.

One domain in which temporal neglect has been extensively studied is in affectively characterizing prolonged experiences. Kahneman (2000) argued that people engage in evaluation by moments—affectively characterizing individual moments instead of summing across them when creating a summary assessment. In evaluating the experience of watching aversive film clips (Fredrickson & Kahneman, 1993), undergoing a painful colonoscopy (Redelmeier & Kahneman, 1996), or even prospectively considering a hypothetical experience from a description of its time course (Varey & Kahneman, 1992; Y. Wang et al., 2023), people seem relatively insensitive to the length of those experiences. Instead, they focus on their experience during a few focal moments (e.g., the most intense and the final moment). In the present context, we examine how people characterize constant rates of change, meaning our own focus is not on identifying which focal moments may capture attention and thus disproportionately inform summary assessments. But if monitoring change is not done continuously, but instead in discrete intervals, the same evaluation-by-moments logic suggests that perceptions of the speed of progress will be sensitive to just how much progress can be gleaned at individual moments (i.e., check-ins) without an appreciation for the underlying time that is passing.

When previous research has examined how repeated sampling of information can distort judgment, it has typically considered how the *nonindependence* of such information can throw things awry. Fiedler (2012) suggested that when people sample information, their judgments fail to take into account the degree of diagnosticity that each additional piece of information carries. Such a failure reflects *metacognitive myopia* (see also Fiedler et al., 2018). As Fiedler et al. (2002) noted, repeated sampling of information is beneficial when each draw is independent (A teacher gets a more accurate perception of their students' satisfaction by reading more students' course evaluations.) But repeated sampling of information rarely produces independent draws, and thus it frequently yields redundant information.

Consider more generally how one might arrive at various conclusions. To know whether an opinion is popular, listen for it being expressed. To know whether a business is thriving, be on the lookout for relevant news reports. Problems arise when one hears a statement repeatedly simply because it is being continually espoused by one opinionated person, or when the positive news coverage one keeps hearing actually reflects many media outlets' reporting on the same event. Perceivers often act as if this actually redundant information is new and informative (Weaver et al., 2007), even after such redundancy is made salient to the perceiver (Begg et al., 1992; Fiedler et al., 2018; Hasher et al., 1977; Unkelbach et al., 2007).

Other research has examined how nonindependence thwarts reasonable inferences from information that—like the focus of the current work—describes a target's progress. That work—much like that of Fiedler and colleagues described above—also considered how perceivers are not sensitive to (partial) redundancies that many descriptions of progress include (Alves et al., 2023; Alves & Mata, 2019; Grüning et al., 2023). Consider watching two basketball games in which one team ekes out a victory by 2 points. In one game, the winning team led the whole game. But in the other game, the two teams traded the leading position throughout the game. If you are like Alves and colleagues' participants, you may be more impressed by the former victor. But note that holding a steady lead is merely a commentary on the early-game performance, because only the opening-minutes' baskets factor into the score throughout the game. Points in the final quarter are reflected in the game score only for the final quarter. Absent a (nonobvious) theory of why early performance is more diagnostic than late performance, attention to who led for the majority of the game reflects a counternormative weighting of redundant information.

Our own focus of study considers how tracking or monitoring progress can lead to counternormative observations, but not due to an inappropriate reliance on redundant information. Instead, we recognize that—much as did the factory supervisor who occasionally dropped by to assess a worker's incremental progress—progress is often tracked by noting changes in output or standing. In this way, our own MFE—like Alves and Mata's (2019) cumulative redundancy bias—suggests that people focus and disproportionately weight a salient index of performance. But whereas Alves and Mata found that people fail to fully appreciate the significance of a global marker of progress (i.e., that repeatedly checking in on cumulative progress gives disproportionate weight to early performance), we instead consider how people fail to properly interpret a local marker of progress (i.e., that how much has been accomplished since a previous check-in depends not only on the amount of change that is observed but on how much time has passed).

When previous researchers have considered the frequency of monitoring, they have primarily focused on how often people track *goal* progress—that is, whether a target (typically, the self) is progressing efficiently toward a desired end state. Furthermore, this research has focused more on determinants of monitoring frequency as opposed to consequences of it (Chang, Webb, & Benn, 2017; Chang, Webb, Benn, & Reynolds, 2017; Chang, Webb, Benn, & Stride, 2017; Webb et al., 2013). For example, people often avoid tracking goal progress, opting to bury their heads in the sand rather than keep close track on how they are faring. Webb et al. (2013) term this "the ostrich problem." When consequences of monitoring frequency have been studied, it has been in terms of actual performance (Jenkins & Terjeson, 2011). Our own hypotheses need not apply to the context of goal pursuit in particular. And although we will identify certain antecedents of monitoring, our focus is on surreptitious monitoring's consequences for perceptions of progress, even when actual progress does not change.

That said, the question of what factors influence monitoring frequency is also relevant for understanding the implications of the MFE. More generally, we suggest that beliefs or expectations about different targets are likely to influence how frequently they are monitored. In that way, the MFE can prove to be a novel mechanism through which prior beliefs or expectations can affect perceptions of others. Previous research on self-fulfilling prophecies (e.g., Carlana, 2019; Figlio, 2005; Gentrup et al., 2020; Hill & Jones, 2021; Papageorge et al., 2020) and confirmation bias (e.g., Vedejová & Čavojová, 2021) have identified different ways by which expectations may influence perceptions. Research on the confirmation bias, for instance, has demonstrated how beliefs and expectations can change what information people seek out and what information they neglect (Klayman & Ha, 1987). The MFE proposes a new way in which these beliefs and expectations can color target evaluations even when target progress is held constant, and the actual, objective meaning of the information to which perceivers are exposed is equivalent.

Overview of Studies

We present 10 studies that, collectively, document the MFE, establish the robustness of and predictable variability in the size of the effect, identify (sometimes insidious or counterproductive) preferences for how often people wish to monitor and be monitored, and distinguish among mechanisms that explain why monitoring frequency influences perceptions of progress. Studies 1a and 1b test the MFE by manipulating the frequency with which workers are monitored. Studies 2a and 2b extend on this paradigm to identify factors that encourage a preference for monitoring or being monitored more frequently. Study 3 attempts to replicate the MFE in a nonsocial domain: the spread of a contagious disease. Studies 4 and 5 examine whether the MFE is robust to financial incentives that reward accuracy in forecasting and decision-making. Study 6—in combination with a follow-up study—tests whether the MFE is robust to various changes in how information about the changing attribute is presented. Study 7 varies whether check-ins are actually completed and experienced over time or the output of such monitoring is merely summarized at a single point in time. Study 8 distinguishes two mechanistic accounts of the MFE, whether it emerges due to perceivers' failure to track information about targets or to make use of the temporal information that they do recall.

Study 1a

Both Studies 1a and 1b offered initial tests of the MFE. Participants completed manager simulation tasks in which they monitored four employees. In both studies, two of the employees were equally productive, but one was monitored *rarely* (every 3 days), while the other was monitored *frequently* (every day). We expected that participants would see the rarely monitored worker as more productive than the frequently monitored one.

Method

Participants

One hundred thirty-two undergraduates (88% female, 11% male, 1% who chose to not disclose; $M_{\text{age}} = 20.59$, $SD_{\text{age}} = 5.13$) at the University of Lisbon participated in exchange for course credit.

Procedure

Participants took part in a workplace simulation. They considered being the manager of a factory that made watch parts. Each day, a (supposedly random) selection of employees was checked in on, at which point the participant would learn each target's output since the last time that employee was monitored. After the monitoring period, participants evaluated how productive each employee was. Crucially, the employees were said not to be aware of these inspections; thus, their productivity could not have been affected by them.

Overall, participants monitored four distinct employees during a simulated 10-day period. As participants proceeded through each day of the simulation, they received feedback about how many watch parts some or all of the employees had completed since the last check-in. To avoid ambiguity, each time feedback was provided, it was explicitly labeled as reflecting the employee's progress since the last check-in. During each check-in, the employee's face, as well as the number of watch parts completed since the last check-in, were presented on screen for 4 s. We counterbalanced which White male face—all drawn from the Chicago Face Database (Ma et al., 2015)—was used to represent each employee. Table 1 presents the check-in schedule, as well as how many incremental watch parts each employee was shown to have completed at each check-in.

Table 1
Check-In Schedule and Observed Output at Each Check-In (Study 1a)

	Day										
Employee	1	2	3	4	5	6	7	8	9	10	Total
F	5	6	5	4	6	5	4	6	4	6	51
R	5	—	—	15	—	—	15	—	—	16	51
L	6	4	—	10	5	—	9	6	—	10	50
H	5	—	10	6	—	11	5	—	9	6	52

Note. On days an employee was not monitored, a dash has been inserted. F = frequent; R = rare; H = high; L = low.

The critical targets were employee F(requent) and employee R(are). These two employees were equally productive; they varied only in how frequently they were monitored—frequently (F) or rarely (R). We also included two additional employees: H(igh) and L(ow). These employees were monitored moderately frequently; one (H) had slightly *higher* productivity than the two critical targets, whereas the other (L) had slightly *lower*.

Following the 10 days of monitoring, participants completed a two-item measure of perceived change (in production) about each of the four targets. The targets were identified both by their employee letter and their photo. One item directly asked participants to *rate* each employee's productivity: "How would you rate each employee's productivity?," anchored at 1 (*very low*) and 9 (*very high*). The second item asked participants to make an *estimate*: "If you checked in on each employee on the next day, how many parts do you think they would have completed?" The order of the items was counterbalanced. We standardized and averaged these items to create a two-item *perceived change* (in production) composite ($r = .21$, $p < .001$).

Results

In order to determine whether and how the employees were perceived to differ in their productivity, we conducted a mixed-model analysis. *Employee* was a categorical predictor that identified which of the four employees (F, R, H, or L) was being judged. We included random effects of participant and target face. These account for the nonindependence of trials completed by the same participant or that concern the same face. A significant effect of employee emerged, $F(3, 390.68) = 15.63$, $p < .001$, which indicated systematic heterogeneity in the productivity perceptions (see Table 2). Suggesting that participants were sensitive to actual differences in productivity, the high-productivity worker (H) was judged to be more productive than the low-productivity worker (L), $t(389.03) = 2.14$, $p = .033$. Providing direct evidence of the MFE, the rarely monitored employee (R) was judged to be much more productive than the frequently monitored one (F), $t(391.03) = 6.48$, $p < .001$. It is notable that this (larger) gap emerged despite no actual difference in these employees' productivity.

Study 1b

Study 1b used a similar paradigm to that used in Study 1a. But in this case, we systematically varied whether each of the four targets was monitored rarely or frequently as well as whether they showed high or low output at each check-in. We carefully operationalized these levels so that the rare-high and the frequent-low targets were equally productive. Comparing whether the rare-high target was perceived as more productive than the frequent-low target offered a direct test of the MFE.

In this design, each of these targets had a matching counterpart who displayed (on average) the same output at each check-in. All that differed was whether those check-ins occurred frequently or rarely. In these pairs, the frequently monitored employee was actually three times as productive as the rarely monitored one. But

Table 2
Average Judgments (and Standard Deviations) for Each Measure, by Target (Study 1a)

Measure	Employee F	Employee R	Employee L	Employee H
Rating	5.81 (1.99)$_b$	6.83 (1.57)$_a$	6.09 (1.54)$_b$	6.48 (1.37)$_a$
Estimate	9.09 (11.91)$_c$	13.05 (16.35)$_a$	10.08 (12.74)$_{b,c}$	10.92 (14.45)$_b$
Perceived change	−0.21 (0.78)$_c$	0.24 (0.85)$_a$	−0.09 (0.68)$_c$	0.06 (0.72)$_b$

Note. Perceived change (italicized, to reflect it is a composite) is the average of the standardized measures: The rating of target productivity and the estimate of how many parts they would complete in an upcoming period. Means within each row that do not share a subscript differ at the $p < .05$ level. F = frequent; R = rare; H = high; L = low.

comparing the perceived productivity of these two employees would allow us to assess just how much temporal neglect perceivers displayed. The hypotheses, methods, sample size, exclusion criteria, and analysis plan of Study 1b were preregistered (https://aspredicted.org/TSM_D33).

Method

Participants and Design

One hundred forty-five CloudResearch-approved Americans were recruited from Amazon's Mechanical Turk (AMT). Forty-seven participants failed to pass at least one of two memory-based attention checks (see Supplemental Materials). Per our preregistered criteria, we excluded these participants from analyses, resulting in a final sample of 98 (57% female, 43% male; M_{age} = 44.18, SD_{age} = 13.77). The design was a 2 (monitoring frequency: rare or frequent) × 2 (average output: low or high), fully within-participants.

Procedure

Participants again took part in a workplace simulation, though the design differed from that of Study 1a in six respects. First, participants were told the employees completed widgets (instead of watch parts). Second, we went to greater lengths to reinforce that the check-ins were done surreptitiously. More specifically, we described how the output box of each employee was accessible behind a wall such that employees could not see when (or that) the supervisor had checked in on their progress. On the one hand, surreptitious monitoring is an increasingly common capability in the modern digital workplace. But we included this feature for a different reason. Had the monitoring not been covert, perceivers might assume that a frequently (vs. a rarely) monitored employee was likely spurred by this constant surveillance to put in more effort, meaning their output should not be taken at face value (but instead as an artificially strong signal of their actual productivity). Our design, even if somewhat contrived in this setting, discourages the operation of this additional mechanism and thus allows us to conduct a more conservative test of the MFE.

Third, we made explicit that whether a particular employee was checked in on a given day was a function of factors unrelated to the employees themselves. That is, we wanted to make sure that there was no information conveyed by which employees were monitored more or less frequently. As before, each time feedback on employees' output was provided, it was explicitly labeled as reflecting the employee's progress since the last time they were checked in on. But as a fourth change, before the simulation began, we tested participants' comprehension of this key feature. Regardless of whether participants displayed knowledge of or confusion on this point, we reinforced this central information. Furthermore, all participants were quizzed on this detail again at the study's conclusion. It was one of two attention checks that, per our preregistered exclusion criteria, participants had to answer correctly in order to be included in our analyses.

Fifth, the set of four employees varied more strikingly in their actual productivity. This was because we systematically varied whether each employee was monitored *frequently* (every day) or *rarely* (every 3 days) during a 12-day simulation. Furthermore, we varied whether the average output observed at each check-in was relatively low (four widgets) or relatively high (12 widgets). We chose these values in particular so that we still would have two employees (frequent-low and rare-high) who were equally productive (see Table 3). The frequent-high employee was very productive (three times more so than the focal pair), and the rare-low was very unproductive (one third as much as the focal pair).

Sixth, the key dependent measures were slightly updated. The *rating* item read "Based on what you have seen, how productive

Table 3
Check-In Schedule and Observed Output at Each Check-In (Study 1b)

Employee	Day 1	2	3	4	5	6	7	8	9	10	11	12	Total
Frequent-high	13	12	12	11	13	10	14	11	14	12	10	12	144
Rare-high	—	—	12	—	—	10	—	—	14	—	—	12	48
Frequent-low	5	3	4	3	3	4	5	5	4	4	3	5	48
Rare-low	—	—	4	—	—	3	—	—	4	—	—	5	16

Note. On days an employee was not monitored, a dash has been inserted.

would you say each worker is?" Participants responded on a 9-point scale anchored at 1 (*not at all*) and 9 (*a lot*). The *estimate* item read "In the next two days, how many widgets do you think each of the workers will complete in total (across the two days)?" We selected 2 days because it was equally discrepant (by 1 day) from how often the frequently and rarely monitored employees were monitored. The two items were presented in a counterbalanced order. Per our preregistration, we standardized and averaged these items to create a two-item *perceived change* (in production) composite (r = .79, p < .001).

Results and Discussion

Following our preregistered analysis plan, we began by testing whether participants' productivity judgments displayed normatively appropriate sensitivity to our two manipulations. We submitted the perceived change composite to a 2 (monitoring frequency: rare or frequent) × 2 (average output: low or high) repeated-measures analysis of variance. With monitoring frequency controlled, participants judged the high-average output targets as more productive than the low-average output targets, $F(1, 97) = 526.57$, $p < .001$, $\eta_p^2 = .84$. And with average output (reported at each check-in) controlled, participants judged the (actually more productive) frequently monitored target as more productive than the rarely monitored target, $F(1, 97) = 66.43$, $p < .001$, $\eta_p^2 = .41$. These main effects were qualified by an interaction, $F(1, 97) = 111.19$, $p < .001$, $\eta_p^2 = .53$, which is itself normatively appropriate (because the productivity implications of average output are more differentiating when that output was monitored more frequently vs. rarely).

Even though participants showed this evidence of being sensitive to actual differences in productivity, they also showed evidence of a counternormative MFE. As can be seen in Table 4, the rare-high employee was judged to be more productive than the frequent-low employee, paired $t(97) = 9.91$, $p < .001$, $d = 1.00$, even though the two employees were equally productive. This replication of the MFE was our central preregistered prediction. We proceeded to conduct exploratory tests by comparing the pairs of targets who were either matched on their monitoring frequency (but differed in the average observed output) or matched on their average output at each check-in (but who varied in their frequency of being monitored).

When two targets were equivalent in how frequently they were monitored, participants easily perceived differences in their actual productivity. That is, the high-output frequently monitored target was seen as more productive than the low-output frequently monitored target, paired $t(97) = 30.63$, $p < .001$, $d = 3.10$. Similarly, the high-output rarely monitored target was seen as more productive than the low-output rarely monitored target, paired $t(97) = 12.12$, $p < .001$, $d = 1.23$. Providing initial evidence of temporal neglect, this pattern was less robust when we held average output at each check-in constant but compared the frequently and rarely monitored targets. On the one hand, the rare-high employee was accurately judged as less productive than the frequent-high employee, paired $t(97) = 11.77$, $p < .001$, $d = 1.19$. Though notably, the frequent-low employee was judged as no more productive than the rare-low employee, paired $t(97) = 1.38$, $p = .170$, $d = 0.14$. In other words, for this latter pair, how much time had passed between check-ins was not simply relatively neglected, but fully so. To be clear, we do not claim that the MFE is so strong that perceivers neglect the passage of time altogether, but the fact that such full neglect happened in one of two relevant comparisons speaks to the strength of the MFE.

Study 2a

Studies 1a and 1b offered initial evidence of a MFE. In a workplace simulation, participants who monitored an employee's output more frequently saw that employee as producing less quickly. Studies 2a and 2b continue to use this same monitoring context but examine preferences for monitoring frequency. In Study 2a, participants considered being a supervisor who set a monitoring schedule for employees about whom they had some information. In Study 2b, participants considered being an employee and indicated how frequently they would prefer to be monitored. Because supervisors typically have the most control over how frequently they monitor their supervisees, Study 2a has the potential to asses whether the MFE would likely depress performance perceptions of certain types of workers in particular. Study 2b then tests to what extent the MFE is counterintuitive, which may hinder monitoring targets' impression management.

Specifically, Study 2a tested whether there would be systematicity—based on beliefs and expectations about targets—to how frequently people would choose to monitor targets' progress. Participants, playing the role of supervisors (like in the previous studies), considered employees who varied in how important their productivity was, the length of time they had spent on a workplace team, and their reputations as particularly good or bad employees. We expected that supervisors would display a bias toward being particularly vigilant of those engaging in important work, those who were new to the job, and those who were suspected to be slackers. Such differential monitoring preferences would be important—and at times insidious—given the results of Studies 1a and 1b showing a MFE in this context. We first conducted exploratory Supplemental Study A ($N = 55$), which tested

Table 4
Average Judgments (and Standard Deviations) for Each Measure, by Target (Study 1b)

Measure	Employee: Frequent-high	Rare-high	Frequent-low	Rare-low
Rating	8.38 (1.09)$_a$	6.09 (1.89)$_b$	3.51 (1.49)$_c$	3.15 (1.66)$_c$
Estimate	20.07 (5.86)$_a$	13.43 (6.74)$_b$	7.24 (2.61)$_c$	6.91 (4.02)$_c$
Perceived change	1.11 (0.50)$_a$	0.22 (0.76)$_b$	−0.70 (0.38)$_c$	−0.79 (0.53)$_c$

Note. Perceived change (italicized, to reflect it is a composite) is the average of the standardized measures: The rating of target productivity and the estimate of how many parts they would complete in an upcoming period. Means within each row that do not share a subscript differ at the p < .05 level.

this idea with a smaller sample size and before we had implemented several improvements in the current preregistered study. As described in the Supplemental Materials, those results affirmed all of those reported in preregistered Study 2a. The hypotheses, methods, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/9CT_SB5).

Method

Participants

One hundred twenty-four CloudResearch-approved Americans (59% female, 41% male; $M_{\text{age}} = 42.95$, $SD_{\text{age}} = 11.51$) were recruited from AMT. No participants failed our preregistered memory-based attention checks (see Supplemental Materials); thus, all are included in the analyses reported below.

Procedure

Participants considered the same workplace context used in Study 1b. But in this study, instead of receiving information about the work output of employees, participants instead set a monitoring schedule for three pairs of employees. More specifically, participants were told that their "goal is to learn **how productive each employee actually is**, so that you can make appropriate personnel decisions" (bolding in original materials, for added emphasis). Participants learned they had time to check-in on only one employee in each pair on each of the following 12 days. Their task would be to decide how many of those days they wanted to check-in on one employee versus the other.

As in Study 1b, participants were quizzed on what they would observe in each employee's output box during a check-in (how many widgets had been completed since the last check-in, not how many widgets were completed that day). In addition, participants were also quizzed to make sure they understood that the monitoring was done surreptitiously—that is, that "the output box is behind a wall, and thus workers cannot see that or how often you stop by." This was crucial so that participants would internalize that monitoring was merely in the service of information gathering, not motivating the employees. All participants received feedback based on their quiz responses, meaning these critical details of the simulation were reinforced for everyone.

Participants then received blurbs describing the six employees. The employees composing each pair differed along different dimensions: the alleged importance of the employee's productivity, their tenure on the team, and the valence of rumors that spoke to each worker's productivity (see Table 5). The three pairs, as well as the two employees composing each pair, were presented in a counterbalanced order. Participants indicated on how many of the 12 days they wanted to check-in on each employee in each pair. If participants' responses for a pair did not sum to 12, they were prompted to revisit their answers.

Results

For each pair of employees, participants tended to allocate more check-ins to one of the employees over the other. More specifically, participants allocated more check-ins to the employee whose productivity was *important* (I; $M = 7.58$, $SD = 2.57$) instead of the employee whose productivity was relatively *unimportant* (U; $M = 4.42$, $SD = 2.57$), paired $t(123) = 6.86$, $p < .001$, $d = 0.62$. They also allocated more check-ins to the *new* employee (N; $M = 8.42$, $SD = 1.95$) instead of the *long-time* employee (L; $M = 3.58$, $SD = 1.95$), paired $t(123) = 13.82$, $p < .001$, $d = 1.24$. Finally, participants assigned more check-ins to the employee about whom *bad* rumors were swirling (B; $M = 8.06$, $SD = 1.80$) compared to the one about whom there was *positive* gossip (P; $M = 3.94$, $SD = 1.80$), paired $t(123) = 12.75$, $p < .001$, $d = 1.15$.[1]

These results acquire importance in light of the MFE observed in Studies 1a and 1b. Employees whose productivity is particularly important, who are new, and about whom negative rumors are prevalent may all start out at a systematic disadvantage. To the extent employers keep a more watchful eye on these employees, supervisors may not appreciate that it is their own (frequent) monitoring schedules that give rise to a sense that less work is getting done. Study 2b, by testing whether the MFE is counterintuitive, has the potential to bolster this concern.

Study 2b

Study 2b examined monitoring frequency preferences in a complementary way. Monitors typically have control over just how frequently they decide to check-in on a target's progress. But the monitored can also play a role. An employee can try to attract their boss's attention and thus have their work monitored frequently, or they can try to work in the shadows, outside of their boss's vigilant eye. Whereas monitors are most likely to have an information-gathering goal, the monitored themselves will typically have an impression-management goal: a desire to be viewed as a good, productive worker.

In Study 2b, we again described the same workplace context used in the previous studies. Whereas in Study 2a, we placed participants in the role of the supervisor, this time we asked participants to consider being an employee. We asked participants—if their goal were to achieve a positive impression as a productive worker—whether they would prefer to be monitored more or less frequently than their coworkers. As with the previous studies, we specified that employees would never know when, exactly, the supervisor checks in on their output box; thus, there was no possibility to put in extra effort on days in which their work is monitored.

We initially conducted Supplemental Study B ($N = 958$), which provided preliminary evidence that people's monitoring frequency preferences are incongruent with the MFE. In Study 2b, we also measured three beliefs that we thought could explain participants' preferences. Most critically, we directly measured whether participants could anticipate the MFE or whether it was counterintuitive. But we also measured whether participants thought an employee would benefit from being the more frequent object of their supervisor's attention (as more frequent monitoring would entail) as well as whether participants thought more frequent monitoring would lead to more accurate productivity impressions. This would allow us to determine whether a misunderstanding of the MFE might independently predict monitoring frequency preferences above and beyond

[1] Note that because the responses within each pair were constrained to sum to 12, these statistical tests and effect sizes are equivalent to what would be observed from one-sample t tests against 6, which would reflect an equal allocation of check-ins between the two targets.

Table 5
Employee Descriptions (Study 2a)

Dimension	Employee	Description
Importance of productivity	I	"Based on when the employee completes their task, it is particularly important that they be a productive worker."
	U	"Based on when the employee completes their task, it does not matter that much whether or not they are a productive worker."
Time on the team	N	"This employee just joined the team, so you don't yet know a lot about them."
	L	"This employee has worked for your company for many years and you have known them for a long time."
Valence of rumors	B	"You have heard negative rumors that this employee can sometimes be slow and inefficient, which makes you nervous about the employee's productivity."
	P	"You have heard positive rumors that this employee is fast and efficient, which makes you confident about the value of this employee as a team member."

Note. I = important; U = unimportant; N = new employee; L = longtime employee; B = bad rumors were swirling; P = positive gossip.

these other potential contributors. The hypotheses, methods, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/1FG_QVD).

Method

Participants

We requested 846 CloudResearch-approved American participants from AMT. In actuality, we received one more than requested. Per our preregistered exclusion criteria, we excluded 45 participants who did not pass the memory-based attention check (see Supplemental Materials). This left 802 participants (67.7% female, 31.3% male, 0.9% nonbinary, 0.1% genderqueer; $M_{\text{age}} = 42.84$, $SD_{\text{age}} = 13.36$) in all analyses reported below.

Procedure

Participants were offered information about a workplace context that paralleled that used in Studies 1b and 2a. Except in this study, participants considered being one of the four employees instead of the supervisor. They learned that their supervisor would occasionally evaluate the employees' incremental output since their last check-in ("the supervisor sees how many widgets the employee has completed since the last time they monitored that employee"). As in Studies 1b and 2a, we specified that each worker's output box is accessible to the supervisor behind a wall, such that the employees cannot see if and when the supervisor has checked in on them. To further make the context parallel to the one faced by supervisors in the simulations used in Studies 1a and 1b, we explained that the supervisor would occasionally evaluate each employee by "rating how productive the employee is and estimating how many widgets they can complete in a day."

Participants were asked,

> If your goal were to try to convince the supervisor that you are a particularly productive employee, would you want your supervisor to perform check-ins on your progress more or less frequently than they perform check-ins on your fellow employees?

Participants responded on a 7-point scale. The order of the seven response options was counterbalanced, but *monitoring frequency preferences* were always coded in this way: 1 = *much less frequently*, 2 = *somewhat less frequently*, 3 = *a little less frequently*, 4 = *equally frequently*, 5 = *a little more frequently*, 6 = *somewhat more frequently*, 7 = *much more frequently*.

Of course, there could be several reasons—beyond a mere misappreciation of the MFE—why employees might want to be the focus of their supervisor's attention. We thus probed three beliefs that related to monitoring frequency. One item (*MFE*) directly assessed an intuition for the MFE: "If two workers were equally productive, do you think that one whose output box was monitored more frequently, or one whose output box was monitored less frequently, would seem more productive in the eyes of their supervisor?" (1 = *definitely less frequently*, 4 = *both equally/does not matter*, 7 = *definitely more frequently*). A second item (*attention*) assessed the perceived benefits of being a more consistent focus of one's supervisor's attention:

> Do you think employees benefit (i.e., are seen as more of a productive, hard worker) when their supervisors spend more time focusing on them and their work output or when their supervisors spend less time focusing on them and their work?

(1 = *definitely less time*, 4 = *both equally/does not matter*, 7 = *definitely more time*). A third item (*accuracy*) assessed the perceived accuracy advantage of checking-in on an employee more often: "Do you think supervisors gain a more accurate sense of an employee's productivity by monitoring their output box more frequently or less frequently?" (1 = *definitely less frequently*, 4 = *both equally/does not matter*, 7 = *definitely more frequently*). Again, the directionality of the seven response options was counterbalanced.

Results and Discussion

We split our analyses into three parts. First, we assess whether participants would prefer to be monitored more or less frequently than fellow employees. Second, we determine whether there is a systematic directionality to participants' three beliefs, which includes a critical test of whether the MFE is intuitive. Third, we examine whether one or more of these beliefs are independent predictors of preferences for being monitored relatively frequently.

Employee Preferences

To begin, we compared participants' preferences for being monitored more or less frequently than their fellow employees against the no-preference point of the scale (4). Overall, participants had a systematic preference for being monitored more frequently than their fellow employees ($M = 4.57$, $SD = 1.40$), $t(801) = 11.55$, $p < .001$, $d = 0.41$. As can be seen in Table 6, almost three times as many participants preferred to be monitored more frequently (as opposed to less frequently) than their fellow employees.

Lay Beliefs About the Effects of Monitoring Frequency

That participants preferred to be monitored relatively frequently does not necessarily mean that they fail to intuit the MFE, much less that such an intuition explains participants' preference for being monitored frequently. We thus turned to examining participants' three lay beliefs. There was systematicity to each belief. First, participants not only failed to anticipate the MFE but actually thought that being monitored *more* frequently would help to make an employee seem more productive ($M = 4.42$, $SD = 1.72$), $t(801) = 6.99$, $p < .001$, $d = 0.25$.

But participants also possessed other beliefs that might encourage a preference for more frequent monitoring. That is, participants also thought they would benefit from being the focus of their boss's attention, as occurs with more frequent monitoring ($M = 4.17$, $SD = 1.84$), $t(801) = 2.62$, $p = .009$, $d = 0.09$. In addition, participants thought that supervisors would achieve more accurate impressions of their employees' productivity through more frequent monitoring ($M = 4.88$, $SD = 1.68$), $t(801) = 14.92$, $p < .001$, $d = 0.53$. Although these beliefs were all held in the aggregate, Table 6 indicates what percentage of participants held each belief.

Which Lay Beliefs Independently Predict Monitoring Frequency Preference?

In a final step, we asked whether one or more of the lay beliefs predicted participants' monitoring frequency preferences. We regressed the monitoring frequency preferences on all three lay beliefs. In this model, all three lay beliefs independently predicted monitoring frequency preferences (see Table 6). In other words, although a misunderstanding of the MFE helped to explain participants' monitoring frequent preferences, other beliefs—the validity of which are less clear—independently predicted these preferences as well.

Study 3

On June 4, 2021, the U.S. state of Florida became the first in the nation to move from a daily report of new COVID-19 infections to a weekly one. The announcement came from the press secretary of Ron DeSantis, Florida's Republican governor who built a national profile in part due to his opposition to pandemic-related precautionary measures. The governor's office explained that progress against the virus meant that there was no longer a need to issue the daily reports (Paz, 2021). But is a shift to infrequent monitoring a way to reinforce perceptions that the threat posed by COVID-19 is waning?

In Study 3, participants completed a simulation in which they took on the role of a regional epidemiologist. We tested whether infrequently (vs. frequently) monitoring the spread of a contagious disease—in the form of occasional (instead of daily) reports of new infections—may make the disease seem to be progressing through a community more quickly. Whereas Studies 1a and 1b tested for this MFE in a workplace context in which progress was good (i.e., worker productivity), Study 3 tests for the MFE in a medical context in which progress is bad (i.e., disease spread). After all, there is nothing inherent to our theoretical logic that suggests the MFE should be specific to social targets.

We expected to find support for the MFE, such that more infrequent updates on the number of new cases would encourage a sense that the disease was progressing quickly. In so doing, we addressed an alternative explanation for Studies 1a and 1b that more infrequent monitoring simply encourages more positive impressions of targets, and extended the effect beyond a social domain. The hypotheses, design, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/LDH_27X).

Method

Participants

We requested 100 Americans from AMT from CloudResearch's approved participant pool.[2] At the study's conclusion, we asked participants whether they received information about both cities on every day of the simulation. Per our preregistered exclusion criteria, those 22 participants (59% female, 41% male; $M_{age} = 39.69$, $SD_{age} = 9.62$) who did not accurately respond to the memory-based attention check (see Supplemental Materials) were excluded from the analyses.

Procedure

Playing the part of a regional epidemiologist, participants learned that part of their job entailed monitoring the spread of a novel virus in two similarly sized cities, Washington and Franklin. They would receive one or two reports on each day, from one or both cities. The report detailed how many new infections had been reported in that city since the last report. We explained that "the two cities simply have different official policies regarding what days of the week they issue these reports." So that participants would understand there was a direct correspondence between infection rates and hospital demand—a detail that was particularly relevant to one of our two key measures—participants learned that approximately 2.1% of those infected would require hospitalization. We quizzed participants on whether each report would include the number of new infections that had been recorded since the last report (as opposed to on that day in particular). Just before the simulation began, this detail was reinforced to all participants.

The simulation lasted 12 days. At the start of each day, the screen was cleared except for a simple timestamp ("Day X"). Next, one or two new case reports were offered in sequence. For one of the cities, a report was issued every day. For the other city, a report was issued every 3 days. Which city was rarely or frequently monitored was counterbalanced across participants. Despite variation in monitoring frequency, 1,171 citizens of each city were infected over the course of the simulation (Table 7).[3]

[2] We first conducted exploratory Supplemental Study C ($N = 200$), which provided initial support for the basic idea tested here. Based on the large effect observed in that study, we preregistered a smaller sample size for Study 3 and made a number of methodological improvements (see Supplemental Materials).

[3] The values themselves were based on actual COVID-19 new-infection reports released by a large U.S. city just prior to the period when this study was run.

Table 6
Response Distributions of Monitoring Frequency Preferences and Lay Beliefs, and Coefficients of Lay Beliefs (Study 2b)

Measure	Response (%) 1–3	4	5–7	*M* (*SD*)	Predicting: preference *B* (*SE*)
Monitoring frequency preference	16.7	36.2	47.1	4.57 (1.40)***	
MFE	23.1	35.0	41.9	4.42 (1.72)***	0.20 (0.03)***
Attention	32.2	25.8	42.0	4.17 (1.84)**	0.17 (0.03)***
Accuracy	16.5	28.4	55.1	4.88 (1.68)***	0.13 (0.03)***

Note. The percentages under Response (%) describe the percentage of participants who gave an answer in each range. Each mean in the *M* (*SD*) column was tested against the midpoint (4). *B* refers to the multiple regression coefficient (and standard error [*SE*]) of each lay belief as a simultaneous predictor of monitoring frequency preference. MFE = monitoring frequency effect.
** $p < .01$. *** $p < .001$.

After completing the 12 days of monitoring, participants completed two measures designed to probe (mis)perceptions that the disease was progressing more quickly in one community than the other. The *progress* measure read "To what extent is the disease progressing quickly through the population of … ?" Participants judged each city—presented in a counterbalanced order—on a 9-point scale anchored at 1 (*not at all*) and 9 (*extremely*). The *hospital* measure instead asked participants to make a policy recommendation:

> Although both cities have the same number of permanent hospital beds, you have the ability to authorize the supply of temporary hospital beds to one or both cities. Now that you have a sense for how quickly the disease is spreading in each city, which city in your opinion has a greater need for more hospital beds?

The 9-point scale was anchored at 1 (*definitely* [City X]) and 9 (*definitely* [City Y]). The midpoint (5) was labeled "They both have the same need." We counterbalanced which city occupied which endpoint. We recoded all responses so that higher numbers reflected an opinion that the rarely monitored city should be prioritized for more beds.

Results and Discussion

We test whether the MFE emerged even when progress—here, community disease progression—is a negative quality. The progress measure showed that the disease was perceived to be progressing more quickly through the rarely monitored city ($M = 6.53$, $SD = 1.82$) than the frequently monitored city ($M = 5.88$, $SD = 1.91$), paired $t(77) = 2.96$, $p = .004$, $d = 0.33$. The hospital measure—through a direct comparison with the scale midpoint (5)—revealed a preference for diverting hospital resources to the rarely instead of the frequently monitored community ($M = 5.79$, $SD = 1.90$), $t(77) = 3.70$, $p < .001$, $d = 0.42$. Of the 46 (of 78) participants who expressed a preference for supplying one community with more additional hospital beds than another, 35 (76%) of them offered the opinion that the rarely monitored community needed more beds. A binomial test showed that this was significantly greater than 50%, $p < .001$.

These results provide another demonstration of the MFE and argue for its generalizability. Frequent monitoring appears to minimize perceptions of progress not merely with social targets and in domains in which slow progress is a negative (e.g., work output), but also in nonsocial domains and when slow progress is taken as a positive thing (e.g., disease spread). Furthermore, these findings emerged in a societally meaningful context showing that the MFE extends to also affect policy recommendations.

Study 4

To this point, we have replicated the MFE in two different contexts—one social and one nonsocial—and demonstrated its counterintuitiveness. That said, perhaps the MFE emerged only because participants were not incentivized to be accurate. If so, perhaps it is not that the MFE is counterintuitive, but that the MFE might disappear or reverse (as Study 2b participants thought) when monitors have more of a stake in making accurate judgments. For that reason, some participants in Study 4 were financially incentivized to make accurate forecasts.

Furthermore, we varied when (some) participants learned of these incentives. Some learned of them only *postmonitoring* (thereby allowing the incentives to encourage more careful formulation of progress judgments), whereas others learned of them *premonitoring* (so the incentives could encourage more careful tracking of the presented target information as well). Of key interest would be whether the MFE was robust to the presence of incentives and, more subtly, whether the size of the MFE might reduce when incentives were introduced (perhaps as a function of their timing).

Table 7
Monitoring Schedule and New Cases Reported at Each Check-In, by City (Study 3)

City	Day 1	2	3	4	5	6	7	8	9	10	11	12	Total
F	105	104	104	97	88	89	81	90	98	102	102	111	1,171
R	—	—	313	—	—	274	—	—	269	—	—	315	1,171

Note. On days a city report was not received, a dash has been inserted. F = frequently monitored city; R = rarely monitored city.

In Study 4, we moved to another nonsocial context in which to study the MFE: national per capita CO_2 emissions. We selected this context because understanding which countries are contributing to total emissions more versus less quickly is important for understanding how international policies and treaties can be written to address more urgent climate problems. International agencies vary in how frequently they require nations to report on their new emissions. Some (e.g., The European Union) require annual reports, whereas others (e.g., the United Nationals Framework Convention on Climate Change) require less regular updates. The MFE suggests that providing updates on new emissions that cover shorter (annual) timescales may downplay how much a nation seems like a polluter. The hypotheses, design, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/W2W_PGF).

Method

Participants and Design

We requested 300 Americans from AMT from CloudResearch's approved participant pool. We excluded 79 participants who failed to answer correctly either of two preregistered memory-based attention checks (see Supplemental Materials). This resulted in a final sample of 221 (56.1% female, 43.0% male, 0.9% nonbinary; $M_{\text{age}} = 42.60$, $SD_{\text{age}} = 13.44$). The design was a 2 (monitoring frequency: rare or frequent) × 3 (incentive: premonitoring, postmonitoring, control), with the former factor measured within-participants.

Procedure

Participants completed a simulation that lasted for 12 years and that included check-ins of one target frequently (every year) and the other rarely (every 3 years). This simulation used a new context: environmental monitoring of two countries' (Benin's and Ghana's) CO_2 emissions. We chose the two African nations both due to an assumption that our American participants would be relatively unfamiliar with (and thus have few perceived notions about) them, the two countries were almost neighbors (separated narrowly only by Togo), and each country was similar (just a few countries apart in worldwide rankings) in terms of their actual annual per capita emissions output (Ritchie et al., 2023). Although the emissions data to which participants were exposed demonstrated that each nation polluted at the same rate (Table 8), we counterbalanced which country was monitored frequently as opposed to rarely.

Before the simulation began, participants answered a question that tested whether they understood that each report would present the total CO_2 emissions *since* the last report of that country, which would not necessarily be the number of emissions for *that year* in particular. On each trial, participants received a report on one or both countries. To help participants keep track of the targets, they were identified not only by their country name but by their flag as well. After the simulation, participants answered a single question that probed by what amount participants thought total emissions would change in the new 2 years: "How many new CO_2 emissions (in tons per capita) would you guess that each country released in the next 2 years?" (on a sliding scale from 0 to 2, in 0.001-unit intervals). Note that we chose 2 years because it was equidistant between the reporting intervals used for the frequently (1 year) and rarely (3 years) monitored countries.

Unlike in the previous studies, we offered some participants an incentive for accurate responding. This is why we used only an objective measure instead of a subjective one (e.g., on a Likert-type scale) that might ask about each country's rate of polluting. Furthermore, we made sure that all participants knew that the pollution rates they observed during the 12-year simulation were the best information they had when forecasting how the pollution figures were likely to change in the coming 2 years. More specifically, we said: "Following the report, neither country made any changes with regard to environmental controls. In other words, you should assume that their rate of new emissions is likely to continue at the same rate."

Some participants were told that if they could forecast the new CO_2 emissions from both countries within 0.11 tons per capita of the correct answer, they would be entered into a drawing for a bonus payment. To understand whether the introduction of incentives might have an effect only if introduced before the check-ins occurred, we manipulated when this incentive information was (sometimes) provided. Those in the *premonitoring* incentive condition learned about the incentives before the monitoring period began but were also reminded of them before they made their judgments. Participants in the *postmonitoring* incentive condition learned about the incentives for accuracy only after the monitoring, just before they made their forecasts. Those in the *control* condition were given no information about financial incentives.

Results and Discussion

We proceeded to test for the MFE in this new context. We used a mixed model that predicted participants' CO_2 emissions forecasts. We included both monitoring frequency (-1 = frequent, $+1$ = rare) and incentive (categorical) as fixed-effects predictors. To determine whether the MFE was sensitive to the introduction of financial incentives, we allowed the two fixed effects to interact. To account for the nonindependence of participants' two forecasts, we included a random effect of participant. Finally, in case participants universally

Table 8

Check-In Schedule and Emissions (in Tons per Capita) Reported at Each Check-In (Study 4)

	Year												
Country	1	2	3	4	5	6	7	8	9	10	11	12	Total
Frequent	0.593	0.587	0.488	0.438	0.524	0.589	0.482	0.492	0.521	0.623	0.445	0.552	6.334
Rare	—	—	1.669	—	—	1.550	—	—	1.495	—	—	1.620	6.334

Note. On years a country did not release a report, a dash has been inserted. The values of the last three reports from the frequently monitored country were presented in a randomized order.

perceived one of the countries to be more of a polluter than the other, we included a fixed effect of country. This final term did not reach significance, $t < 1$, and thus will not be discussed further.

To begin, we replicated the MFE. Although each country was releasing new CO_2 emissions at the same rate, participants forecast that the rarely monitored country would release more pollution in the next 2 years ($M = 1.32$ tons per capita, $SD = 0.38$) than would the frequently monitored country ($M = 0.70$, $SD = 0.30$), $F(1, 435) = 369.81$, $p < .001$. Crucially, this gap did not vary by the presence of incentives or their timing, $F(2, 435) = 1.41$, $p = .245$. In fact, as can be seen in Table 9, the MFE replicated at each level of the incentive factor. This speaks to the robustness of the MFE.

Study 5

Whereas Study 4 showed that even financially incentivized forecasts showed (similarly sized) evidence of the MFE, Study 5 went a step further in testing whether participants would make a financially disadvantageous decision that the MFE would anticipate. More specifically, participants took part in an investment game in which they had to decide in which of two new ventures to invest their money. Following this decision, participants monitored the performance of both ventures during an initial 12-week simulation that would itself have no direct financial implications for participants themselves. Although participants received weekly reports about the new operating profits of their chosen venture, they received frequent (weekly) or infrequent (every 3 weeks) reports from their unchosen alternative. At that point, and just before each business's returns *would* affect the bonus payout participants might receive, they had the opportunity to switch to the other investment. Though crucially, participants would have to pay a small tax should they wish to switch. Although both businesses' total operating profits (and thus investor returns) were rising at equivalent rates, the MFE predicts that participants would be more likely to make a financially unwise choice and switch to the alternate investment opportunity when its performance progress had been monitored rarely (vs. frequently). The hypotheses, design, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/1MN_Y5D).

Method

Participants and Design

We requested 500 Americans from AMT from CloudResearch's approved participant pool. We received one more participant than requested. Each participant was randomly assigned to one of two monitoring frequency conditions: Participants received reports from an unchosen business opportunity *frequently* or *rarely*. Per our preregistration, we excluded the 54 participants who failed to accurately answer a memory-based attention check (see Supplemental Materials). This resulted in a final sample of 447 (61.3% female, 37.8% male, 0.9% nonbinary; $M_{age} = 40.98$, $SD_{age} = 11.87$) who were included in the analyses.

Procedure

Participants learned they would be taking part in an investment game that would offer them the opportunity to earn actual money based on the savviness of their investment decision-making. We explained:

> In the real world, investors will invest in different ventures, but they are always on the lookout for new opportunities. Sometimes it is wise to stick with one's current investment, whereas in other cases it can make sense to sell one's investments and shift to other opportunities one initially took a pass on.

To begin, participants read information about two businesses that they could invest in: Olympus (a Greek restaurant) and SimplyYoung (a clothing retailer). They had to select one of the two businesses in which to place their initial investment. We explained that this investment would serve as a cash infusion into the company, but that investors would not receive any returns for the first 12 weeks. Starting in Week 13, each of 200 investors (all of whom invested an equal amount) would receive 20% of the operating profits as a return. Because the simulation would continue for 12 more weeks, this meant that participants had a chance to receive 0.5% of the operating profits for Weeks 13–24. To make certain that participants understood that they might actually receive this windfall (and thus to incentivize smart decision-making), we explained that one randomly chosen participant would actually receive the payout as a study bonus.

At this point, participants went through the first 12 weeks of the simulation. For all participants, they received a report every week on their chosen business that indicated by how much total operating profits had grown since the last report. But depending on their condition, they also received reports from the unchosen business that came in *frequently* (every week as well) or only *rarely* (every 3 weeks). As in most of the previous studies, participants were quizzed before the simulation started (and provided with instructions-reinforcing feedback) to make certain they understood that the reports would detail growth in operating profits *since* the last report, not necessarily for that week alone. Furthermore, each new report made explicit that the report detailed new operating profits "since their last report."

Table 9
Monitoring Frequency Effect Across Levels of Incentive (Study 4)

Level of incentive	Country: Frequent	Country: Rare	*B* (*SE*)	*t*(435.00)	*p*
Control	0.72 (0.30)	1.27 (0.37)	0.28 (0.03)	10.67	<.001
Premonitoring	0.69 (0.34)	1.36 (0.41)	0.33 (0.03)	11.33	<.001
Postmonitoring	0.67 (0.25)	1.35 (0.35)	0.34 (0.03)	11.33	<.001

Note. The columns under Country contain the means (and standard deviations) for each country, per level of incentive. *SE* = standard error.

After the initial monitoring period ended (and once actual investor returns were set to begin), participants learned that now that they had more information about how each business was performing, they had an opportunity to switch their investment. In actuality, both businesses were growing their total returns at the same rate (see Table 10).[4] Although all participants were given an opportunity to switch, we explained that those who switched would have to pay a 1% tax on their returns: "If you do wish to switch, your earnings would be taxed at a rate of $0.01 for every $1 of return you receive." This meant that those who stayed with their initial investment had the opportunity to earn a bonus of $41.30, whereas those who changed to the alternate option had the opportunity to receive a bonus of $40.89.

Results and Discussion

We tested whether monitoring a foregone investment opportunity rarely would encourage a financially unwise investment decision, as the MFE would anticipate. When the alternative business opportunity was monitored frequently, only 4.9% of participants elected to switch to it. But when the alternative business opportunity was monitored rarely, the decision to switch skyrocketed to 37.0%, $\chi^2(1) = 75.06$, $p < .001$. In combination with Study 4, Study 5 illustrates that the MFE is robust to financial incentives for accuracy. Study 5 shows that this can occur not merely when directly making forecasts about target attributes' future rates of change, but when making financially incentivized decisions that should depend on those forecasts.

Study 6

Study 6 extended on our previous studies in three ways. First, we moved to a new monitoring context that was inspired by the experience of the second author. Along with the coronavirus pandemic's many other effects, one consequence was a near elimination of his family's weekly get-togethers. As sometimes months would pass between visits, he was impressed by how much his young nephews were going through a growth spurt. The boys' parents—who saw their children every day—were surprised by this observation. Their uncle wondered if the MFE was to blame. Study 6 thus made use of a simulation in which a young nephew's height was measured at family gatherings that happened frequently or rarely.

Whereas the previous two studies found that the MFE was robust to financial incentives, Study 6—in combination with two follow-up studies (Supplemental Studies D and E)—examined whether the MFE was robust to variants in how change was reported. First, whereas the previous studies all varied monitoring frequency within-participants, these studies all varied it between-participants. Second, whereas the previous studies provided information about incremental changes (e.g., newly produced widgets, new operating profits), Study 6 supplied information about the cumulative total directly (i.e., the new height). In this way, a cumulative index of progress—instead of only an incremental one—was explicitly supplied in both conditions. The follow-up studies, described after the main study, examined variants on this paradigm that both stress-tested the MFE further and gave some insight into the conditions under which the MFE is likely to emerge more or less strongly. The hypotheses, methods, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/WXY_KZV).

Method

Participants and Design

One hundred eighty-six undergraduates at the University of California, Berkeley, took part in the study in exchange for course credit. Participants were randomly assigned to one of two *monitoring frequency* conditions: frequent or rare. Per our preregistered criteria, we excluded three participants who failed a memory-based attention check (see Supplemental Materials). This resulted in a final sample of 183 (60.7% female, 38.8% male, 0.5% who did not disclose; $M_{\text{age}} = 20.83$, $SD_{\text{age}} = 2.52$) included in all analyses reported below.

Procedure

Participants took part in a simulation in which they were monitoring the change in height of a young relative, a nephew. It was explained that during family get-togethers that happened the first weekend of each month that the entire family was available, one ceremonial aspect—replicated by families worldwide—was measuring the height of the children to see how much they had grown. During the 2-year simulation, participants in the frequent monitoring frequency condition received an update on the nephew's height every other month. For those in the rare monitoring frequency condition, these family get-togethers happened every 6 months. Notably, the nephew's height at the start and end of the 2-year simulation were equivalent in both conditions (see Table 11).

At each check-in, participants were reminded of the nephew's last recorded height (much as would happen when families mark a growing child's height on a doorframe) as well as the newly recorded height: "Recall that at the last family gathering your nephew was [height] cm tall. When you measured your nephew's height at this family gathering, you found he was [new height] cm tall."[5] To equate the actual passage of time during each simulation, on each month during which a get-together did not take place, participants were told "The family was not able to get together in [month and year]." At the end of the simulation, all participants made a single estimate of how tall they expected their nephew to be 4 months later: "When you measure your nephew's height in April 2024, what do you think it will be? (*Recall it was 125.8 cm in December 2023.*)" Responses were offered on a slider scale that ranged from 125.8 cm (the last-recorded height) to 131.8 cm and could be expressed to the nearest 10th of a centimeter. Note that a forecasting period of 4 months was chosen because it was

[4] Due to a programming error, the B-Rare profits actually grew at 8 cents per week less than the other three businesses' profits. Although we doubt this was detectable by participants (and it has no effect on the size of the potential bonus), it does sometimes make the prospect of switching to the rarely monitored investment opportunity less appealing, which works against the hypothesis.

[5] Americans are much more likely to record height in terms of feet and inches. Given the somewhat complicated customs of recording fractional heights in inches, which are almost always recorded using fractions instead of decimals, and rarely go beyond specifying ½ of an inch, we used cm to allow for the expression of height with a single number (instead of feet and inches), and because the use of decimals (that allow for more fine-grained and thus realistic descriptions of change) is less odd in this context.

Table 10
Check-In Schedule and Operating Profits Reported at Each Check-In (Study 5)

	Week											
Business	1	2	3	4	5	6	7	8	9	10	11	12
A												
Frequent	\$3,361	\$3,463	\$3,436	\$3,421	\$3,347	\$3,466	\$3,523	\$3,448	\$3,537	\$3,350	\$3,534	\$3,418
Rare	—	—	\$10,261	—	—	\$10,234	—	—	\$10,507	—	—	\$10,302
B												
Frequent	\$3,400	\$3,426	\$3,504	\$3,381	\$3,483	\$3,401	\$3,503	\$3,458	\$3,484	\$3,383	\$3,380	\$3,501
Rare	—	—	\$10,330	—	—	\$10,265	—	—	\$10,444	—	—	\$10,264

Note. Whether a business followed a frequent or rare schedule depended on each participant's condition and investment choice. Which business corresponded to A and B was counterbalanced across participants. On weeks participants did not receive a report from the unchosen business, a dash has been inserted.

equidistant between the monitoring intervals used in the frequent (2 months) and rare (6 months) monitoring frequency conditions.

Results and Discussion

We tested whether the MFE was replicated when check-ins were presented not in terms of how much an attribute had changed since the last check-in, but instead in terms of a new cumulative total. As predicted, participants forecasted that the nephew would be taller when they had monitored his height rarely (M = 128.67 cm, SD = 1.11) as opposed to frequently (M = 128.04 cm, SD = 1.10), $t(181) = 3.89$, $p < .001$, $d = 0.58$. Although this replicates the MFE when change was presented indirectly by providing a new cumulative total (i.e., height), we did also provide reminders of the previously observed cumulative total. We wished to understand whether the MFE would be robust to omitting these reminders, and to gain some understanding of how these features (check-in reports that supplied information about incremental changes directly as opposed to cumulative totals, the presence or absence of reminders of the previous cumulative total) affected the size of the MFE. Toward this end, we conducted a preregistered exact replication of Study 6 using an online sample (Supplemental Study D; N = 335 CloudResearch-approved Americans). We again found evidence of the MFE: $d = 0.71$, $t(333) = 6.51$, $p < .001$.

Although we encourage caution in making cross-study comparisons, we conducted preregistered Supplemental Study E (N = 675 CloudResearch-approved Americans) that tested two variants but using the same online population as Supplemental Study D. When only the nephew's measured heights (but no reminders of the previously observed height) were provided, the MFE emerged but more weakly: $d = 0.24$, $t(671) = 2.39$, $p = .017$. The MFE emerged most strongly when the nephew's height was presented only in terms of incremental change since the last check: $d = 1.57$, $t(671) = 13.31$, $p < .001$.

In summary, the MFE emerged most weakly when there were the greatest memory demands on participants, requiring them to encode not only the most recently presented measurement but also the previously presented one as well. By the evaluation-by-moments logic, this removed one advantage that the information about change possessed: its evaluability at a single moment in time. The MFE emerged most strongly when the math that described change—height at time t minus measured height at time $t - 1$—was calculated for participants, thereby placing the fewest demands on participants to absorb that information. Although these findings offer some hints of when the MFE may be observed more strongly versus weakly, most crucial for the present purposes is that the MFE robustly emerged across all of these variants.

Study 7

In monitoring progress over multiple periods of time, one must track both how much change tends to occur at check-ins as well as how frequent those check-ins are. Whereas progress check-ins involve the explicit presentation of information about attributes, time passes more invisibly. To be clear, in our simulations, we do explicitly track and call attention to the *passage* of time. But those periods tick by as something of background information that is less focal than the change updates offered at each check-in. Furthermore, as our evaluation-by-moments logic highlighted, only output information is interpretable at a single moment in time. Tracking monitoring frequency requires integrating information across multiple moments in time, making it an attribute that perceivers are less likely to make use of. Though by this logic, the temporal neglect that is hypothesized to undergird the MFE should be reduced if we could modify the information presentation so that all relevant attributes were presented at once instead of across many moments. To accomplish this, Study 7 experimentally varies whether monitoring happens through sequential experience (as it does when one is

Table 11
Check-In Schedule and Target Height at Each Check-In (Study 6)

	Check-in												
Monitoring frequency	0	1	2	3	4	5	6	7	8	9	10	11	12
Frequent	108.6	109.9	111.6	112.8	114.3	115.7	117.2	118.5	120.2	121.4	122.9	124.3	125.8
Rare	108.6	—	—	112.8	—	—	117.2	—	—	121.4	—	—	125.8

Note. When a check-in did not occur, a dash has been inserted.

actually monitoring a target across time) or whether all of the information was merely summarized retrospectively at a single moment in time (as might occur when one simply receives a retrospective report).

We returned to the CO_2 emissions context used in Study 4. Whereas some participants received our repeatedly used *sequential* presentation format, others received a new *tabular* presentation format that merely summarized the twelve periods of monitoring in a single table. Whereas the sequential presentation format leaves the attribute updates focal as time passes in the background, the table makes explicit both the attribute information but also the passage of time between check-ins. We predicted that this tabular presentational format would reduce the magnitude of the MFE. The hypotheses, methods, sample size, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/L83_2JN).

Method

Participants and Design

One hundred ninety-nine CloudResearch-approved Americans were recruited from AMT. Participants were randomly assigned to one of two *presentation format* conditions: sequential or tabular. Per our preregistered criteria, we excluded 39 participants who failed to pass either of two memory-based attention checks (see Supplemental Materials). This resulted in a final sample of 160 (56% female, 43% male, 1% nonbinary; $M_{\text{age}} = 43.05$, $SD_{\text{age}} = 13.07$).

Procedure

The monitoring context was identical to that used in Study 4. Having found that incentivizing accuracy did not moderate the MFE, we did not use financial incentives. Participants in the sequential condition received information about the new CO_2 emissions using the same approach as in the previous studies. The check-ins were presented sequentially, allowing participants to experience the passage of time (to scale). In contrast, participants in the new tabular condition were presented all of the information about the 12 years of check-ins at a single moment in a table (see Figure 1). Although participants were required to study the table for at least 35 s, they could spend no more than 95 s looking at it. In both conditions, participants could no longer access this information once the dependent measures appeared.

Participants then answered three questions in a randomized order that assessed how quickly new emissions were seen to be released in each country. One item was the dependent measure from Study 4: a prediction of how many new CO_2 emissions each country would release in the next 2 years. A second item asked "Based on what you have seen, how much is each country contributing to rising CO_2 levels in the atmosphere?" (1 = *not much*, 7 = *very much so*). The third item noted that "nongovernmental organizations can direct funding to countries that need more help in reducing their emissions." Participants then indicated to what extent "each country should be prioritized for such funding" (1 = *not much*, 7 = *a lot*). We standardized and averaged these items to create a three-item *perceived change* (in pollution) composite (α = .70).

Results and Discussion

In order to test the MFE, as well as whether it was moderated by the presentation format, we used a mixed model predicting the perceived change composite. We included several fixed effects: monitoring frequency (−1 = frequent, +1 = rare), presentation format (−1 = sequential, + 1 = tabular) as well as their interaction. In addition, we included a random effect of participant (to account for the nonindependence of their responses). Finally, we included a fixed effect of country (in case there was a tendency to perceive one country or the other as more of a polluter; there was not, $t < 1$).

A main effect of monitoring frequency indicated that, overall, we replicated the MFE ($B = 0.32$, $SE = 0.04$), $t(157) = 9.04$, $p < .001$. Participants thought that the rarely monitored country was contributing to pollution more quickly than the frequently monitored nation. Crucially, the MFE depended on the presentation format ($B = -0.17$, $SE = 0.04$), $t(157) = -4.75$, $p < .001$. We proceeded to decompose the effect of monitoring frequency by presentation format. When the check-ins were experienced sequentially, participants showed evidence of a strong MFE ($B = 0.50$, $SE = 0.05$), $t(157) = 9.42$, $p < .001$. When participants received information about the check-ins at a single point in time, the MFE was cut by more than two thirds ($B = 0.15$, $SE = 0.05$), $t(157) = 3.02$, $p = .003$ (see Table 12).

Two aspects of these findings are of note. First, although the tabular presentation format eliminated most of the MFE, it did not eliminate all of it. This suggests that the MFE is not explained only by the temporal neglect encouraged by the relative invisibility of the passage of time. The remainder may be attributable to the simple

Figure 1

Presentation of the CO_2 Emissions in the Tabular Condition (Study 7)

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Ghana	**+0.593** since last report	**+0.587** since last report	**+0.488** since last report	**+0.438** since last report	**+0.524** since last report	**+0.589** since last report	**+0.482** since last report	**+0.492** since last report	**+0.521** since last report	**+0.623** since last report	**+0.445** since last report	**+0.552** since last report
Benin	no report released	no report released	**+1.668** since last report	no report released	no report released	**+1.550** since last report	no report released	no report released	**+1.495** since last report	no report released	no report released	**+1.620** since last report

Note. The order of the last three reports, in the frequent condition, was randomized for each participant. CO_2 = carbon dioxide.

Table 12

Average Judgments (and Standard Deviations) for Each Measure, by Presentation Format and Monitoring Frequency (Study 7)

Measure	Sequential format		Tabular format	
	Frequent monitoring	Rare monitoring	Frequent monitoring	Rare monitoring
New emissions	0.73 (0.34)$_a$	1.35 (0.37)$_d$	1.00 (0.50)$_b$	1.15 (0.44)$_c$
Contributing	3.40 (1.45)$_a$	4.72 (1.66)$_c$	4.19 (1.49)$_b$	4.64 (1.73)$_c$
Priority	3.79 (1.46)$_a$	5.21 (1.59)$_c$	4.35 (1.67)$_b$	4.93 (1.67)$_c$
Perceived change	−0.55 (0.59)$_a$	0.43 (0.68)$_d$	−0.10 (0.73)$_b$	0.21 (0.79)$_c$

Note. Perceived change (italicized, to reflect it is a composite) is the average of the standardized measures: Estimates of how many *new emissions* each target will release in an upcoming period, ratings of how much each target is *contributing* to rising CO_2 levels in the atmosphere, and ratings of how much each target should be a *priority* with funding to reduce CO_2 emissions. Means within each row that do not share a subscript differ at the $p < .05$ level. CO_2 = carbon dioxide.

association between the large observed changes (as the rarely monitored target offers with both presentational formats) and a sense of faster change overall. In other words, the tight connection between the amount of change in the focal attribute and its perceived rate of change persists regardless of the presentation format.

Second, the tabular presentation may not only have made the passage of time more directly visible, but may also have made it easier to encode information about the targets. On the one hand, this could explain why tabular presentation participants were more accurate in their perceptions (given the two countries were shown to actually pollute at the same rate). On the other hand, one might have expected that a difficulty with tracking information in the experiential condition would lead the two targets to be blurred (and thus seen as especially similar); this would have worked against the MFE. The final study will lean only on the repeatedly used sequential presentation format and test whether the MFE can be explained by biases in memory for the tracked information (which would lend credence to this alternative explanation for why the tabular presentation reduced the MFE) or instead to a failure to properly synthesize the target information as it is recalled (consistent with the idea that the passage of time, as an experienced variable, may be relatively neglected when forming impressions of rates of change).

Study 8

We have provided support for the MFE across various contexts. Those studies also emphasized the robustness of the effect while also identifying factors that contributed to predictable variability in the size of the MFE. By our reasoning, people display a sort of temporal neglect: They are more sensitive to the greater incremental progress that infrequent monitoring begets without fully adjusting for the background passage of time. That said, there remains ambiguity regarding exactly what form this temporal neglect takes. One possibility is that output (or change) information is itself tracked more accurately than the temporal information. A second possibility is that the target information (as recalled) is not synthesized normatively to fully incorporate the implications of the different temporal frequency of monitoring.

By the first account, the challenge with monitoring progress may simply be that people are less adept at accurately tracking (and thus accurately distinguishing between targets with regard to) how often a target is monitored. That is, given incremental progress is directly observed, it may be simpler to distinguish the (greater) incremental progress of a rarely (from a frequently) monitored target. Keeping up with how frequently monitoring is occurring at all—given each check-in is experienced less by its frequency and more by the salient output information that is presented during such check-ins—may be a more difficult tracking task. Under this possibility, the MFE would emerge due to a tendency to differentiate targets more based on how much incremental output they displayed than on the different frequency with which they were monitored, even though perceivers then lean on this tracked (but biasedly recalled) information in a normative manner when drawing conclusions about target productivity. That said, we were a priori skeptical of this account's plausibility given previous work suggesting that people can automatically track the frequency of events (Hasher & Zacks, 1984). In this sense, it seemed unlikely that monitoring frequency itself would be especially plagued by the sort of poor tracking that could give rise to the MFE.

Note that the second account—that perceivers would relatively neglect the temporal component when forming productivity impressions—is most consistent with the logic developed in the introduction, particularly the idea of evaluation by moments (and the temporal neglect that implies) that seems to characterize people's evaluations of affective experiences (Fredrickson & Kahneman, 1993; Kahneman, 2000; Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Varey & Kahneman, 1992). Of course, the MFE is not about affective evaluations. But just as people neglect the duration of affective experiences and instead prioritize the intense affective experiences of single focal moments, the MFE itself may reflect a more robust reliance on the feedback gleaned from single moments (i.e., observed output) without a full consideration of how monitoring frequency should also guide interpretation of these data points. Note that the tabular condition in Study 7 essentially consolidated the full period of tracking into a single moment, which should have worked against the tendency to neglect monitoring frequency (which otherwise occurs across moments). Study 8 more directly tests whether perceptions of progress show greater reliance on recollections of observed output than monitoring frequency.

Study 8 sought to test these two accounts by returning to the worker monitoring paradigm used in Studies 1a–2b. Notably, these two accounts are not mutually exclusive; each may contribute to the MFE. In addition to a few other modifications, we added measures that asked participants to recall how much incremental output each target tended to display at each check-in and how often those check-ins occurred. This allowed us to replicate the MFE and test whether biased recall about the targets and/or relative neglect of the temporal information (as recalled) contribute to the MFE. The materials, sample size, hypotheses, exclusion criteria, and analysis plan were preregistered (https://aspredicted.org/83J_16g).

Method

Participants

Three hundred seventy-two Americans were recruited from AMT. Per our preregistered exclusion criterion, the 97 participants who failed to answer correctly either or both memory-based attention checks (see Supplemental Materials), were excluded from all analyses. This left a total of 275 participants (70% female, 28% male, 1% nonbinary, 1% genderqueer; $M_{\text{age}} = 38.77$, $SD_{\text{age}} = 12.26$) in the analyses reported below.

Procedure

The procedure took a similar form to that of Studies 1a and 1b, but with the following changes. First, we included only two employees (the rarely and frequently monitored with the same total output; see Table 13), and thus omitted the other two employees used in Studies 1a and 1b. Second, and most critically, we expanded our list of measures to allow us to distinguish between the mechanistic accounts.

Perceived Productivity. We retained both items—asking participants to *rate* and *estimate* how productive the employee is—from Study 1b. Note that, for the estimate item, we asked about 2 days because it was equally different from the interval at which participants monitored the two targets (i.e., 1 and 3 days). Participants first completed one item for both targets before proceeding to the other. We counterbalanced the order of the items and the order of the targets within each item. To create a single index of the perceived pace of change (in work output), we standardized each item and averaged them ($r = .37$, $p < .001$) to create a two-item *perceived productivity* composite.

Recalled Target Information. Next, we had participants report, for each target, how many completed parts they tended to observe each time they surveyed the output box and how many times the target was monitored. The two items were presented in a counterbalanced order, and the order of the targets within each item was counterbalanced as well.

Recalled Observed Output. For each target, participants responded to "On the days you monitored the employee, how many parts (on average) did you find in their box?" The actual averages were 4 and 12. In an effort to avoid the contaminating influence of outlier responses, we imposed a constraint that participants had to answer between 0 and 20 parts, inclusive.

Recalled Monitoring Frequency. For each target, participants were asked "On how many days, out of the 12, did you monitor each employee?" Here, we imposed the constraint that responses had to be between 0 and 12, inclusive.[6]

Results and Discussion

We begin by testing whether the MFE replicated in this somewhat modified paradigm. At that point, we move to our new measures in an effort to test the potential mechanistic accounts:

Monitoring Frequency Effect

In order to test for the MFE, we defined a mixed model predicting the perceived productivity composite with monitoring frequency (+1: rare, −1: frequent) as a fixed-effects predictor. Given each participant offered these judgments about two targets, we included a random effect of participant. A significant effect of monitoring frequency—$B = 0.28$, $SE = 0.03$, $t(548) = 8.35$, $p < .001$—showed that the rarely monitored target was seen as more productive than the frequently monitored target. The untransformed means and between-target comparisons by item are provided in Table 14.

Should the MFE Have Followed From the Target Details as Recalled?

For each target, we multiplied how many times the participant remembered checking-in on a target by the average amount of output that they remembered observing at each of those check-ins. This product reflects a *normative productivity calculation*, an index of how productive a worker would be seen to be if participants optimally synthesized their own recollections about both observed output and monitoring frequency into a productivity perception. Using this index, the rarely monitored worker should not have been seen as any more productive ($M = 49.09$, $SD = 27.68$) than the frequently monitored worker ($M = 48.41$, $SD = 22.59$), $B = 0.34$, $SE = 1.08$, $t(548) = 0.31$, $p = .754$. This clearly contradicts the first mechanistic account; the MFE cannot be traced to recall mistakes.

To be clear, the fact that the normative productivity calculation does not differ between the targets does not mean that participants tracked the output and monitoring frequency variables accurately. In fact, as can be seen in Table 15, the actual recall for each attribute for each target was significantly biased toward the overall mean. But this bias in recall was essentially identical in magnitude for each attribute. If memory biases were to explain the MFE, this bias in recall would have needed to be greater for the recalled monitoring frequency than the recalled output attribute. As a formal test of the recall errors' statistical distinguishability, for each participant and for each recalled feature, we took the difference score in the recollections for the two targets and divided that by the actual difference. In this way, a quotient of 1 would indicate unbiased (relative) recall. Although both quotients were less than 1 (Ms = 0.63 and 0.64), they did not significantly differ, $t < 1$. The near equivalence of this bias helps to explain why the product of the two variables (i.e., the normative productivity calculation) did not differ between the targets.

Were the Recalled Target Details Normatively Applied?

If the MFE cannot be traced to bias in how the observed output and monitoring frequency were recalled, then this suggests it may be *how* these target recollections were leaned upon that may explain the MFE. In particular, we tested our second account by examining whether perceived change was more a function of recalled observed output than recalled monitoring frequency. Note that as each of these (recalled) variables increases—holding the other one constant—we should see an accompanying increase in perceived change. But if participants display the evaluation-by-moments approach that is

[6] Because the frequently monitored target was monitored every day, responses could not err in a positive direction. This essentially guaranteed that we would observe significant negative bias in such recall.

Table 13
Check-In Schedule and Observed Output at Each Check-In (Study 8)

Target	Day 1	2	3	4	5	6	7	8	9	10	11	12	Total
F	5	3	4	3	3	4	5	5	4	4	3	5	48
R	—	—	12	—	—	10	—	—	14	—	—	12	48

Note. On days an employee was not monitored, a dash has been inserted. F = frequent; R = rare.

core to our second mechanistic account, then we should see bias in which of these attributes guides productivity judgments.

To test this second account, we entered the two recalled variables—recalled observed output (per check-in) and recalled monitoring frequency—as fixed effects into the model that already included employee as a predictor of the perceived change composite. This model showed that as recalled observed output increased—holding recalled monitoring frequency constant—perceived change increased (B = 0.24, SE = 0.04), $t(546) = 5.26$, $p < .001$. But as recalled monitoring frequency increased—holding recalled observed output constant—there was no tendency for perceived change to go up (B = 0.07, SE = 0.05), $t(546) = 1.50$, $p = .135$. A test of whether these betas were different in magnitude was significant, $t(546) = 2.65$, $p = .008$.

Transparency and Openness

With the exception of Study 1a, all studies in the main article were preregistered and links to the preregistrations are provided. The preregistrations include the methods, sample sizes, and analysis plans. We describe our designs and report all of the analyses preregistered. All materials, data, and analysis code can be accessed online (https://osf.io/kt8rj/?view_only=3722dd9d1edd442aab4fc2b24a20de27).

General Discussion

When tracking change, perceivers do not merely observe the final result of a progression. Instead, they get updates along the way, which offer a sense of whether a target attribute is rapidly changing or relatively stagnating. When a situation is in rapid flux, this usually calls for more consistent monitoring. And when in a period of relative stagnancy, there is usually little motivation to expend the time and effort to assess change. As such, one might expect that perceivers anticipate and come to see more change when engaging in more frequent monitoring.

Table 14
Effect of Monitoring Frequency on the Separate Measures (Study 8)

Measure	Monitoring frequency: Frequent	Rare	*B* (*SE*)	*t*	*p*
Rating	6.32 (2.05)	7.21 (1.74)	0.45 (0.08)	5.50	<.001
Estimate	7.38 (2.75)	10.13 (4.90)	1.38 (0.14)	9.61	<.001
Perceived change	−0.28 (0.69)	0.28 (0.86)	0.28 (0.03)	8.35	<.001

Note. Perceived change (italicized, to reflect it is a composite) is the average of the standardized measures. In the Monitoring frequency columns, values are means (and standard deviations). The inferential statistics, with betas and standard errors (*SE*s), describe a test of the effect of monitoring frequency in that row.

This intuitive hypothesis misses a key consequence of monitoring frequency: Depending on whether one maintains a watchful eye or a more inattentive one, how much change will be observed at each check-in will vary, independent of any actual differences in the rate of progress. Across a variety of monitoring contexts—workplace productivity (Studies 1a–2b, and 8), disease spread (Study 3), environmental emissions (Studies 4 and 7), business returns (Study 5), and human growth (Study 6)—we found that change seemed to occur more slowly when a target was monitored more frequently.

Because different beliefs or expectancies are likely to mold perceivers' monitoring frequency preferences (Study 2a), these causal precursors are likely to have predictable and sometimes insidious consequences on judgments of progress. Further complicating matters, the MFE is counterintuitive, which can lead to potentially counterproductive preferences regarding how frequently one is monitored (Study 2b). The MFE also anticipated misguided preferences for how public resources would be allocated (Study 3) and where investments were made (Study 5). Even when participants were financially incentivized to make accurate forecasts (Study 4) or decisions (Study 5), monitoring frequency was varied between- instead of within-participants (Study 6), and changes in attribute values had to be calculated by participants instead of directly supplied by the experimenter (Study 6, Supplemental Studies D and E), the MFE remained.

We appealed to a combination of relevance insensitivity (that people are often immune to background features that change how focal attributes should be interpreted), evaluation by moments (by extending the idea from summarizing prolonged affective experiences to prolonged attribute tracking), and temporal neglect to explain the MFE. When participants no longer had to engage in monitoring across time but instead had all check-in information summarized for them at a single moment in time, the MFE was reduced (Study 7). Although participants displayed some memory biases for the information they tracked, these were no greater for the temporal information (the amount of time that passed between check-ins) than the observed output at each check-in; it thus could not account for the MFE. Instead, participants failed to normatively synthesize the information as recalled, displaying temporal neglect of the passage of time between check-ins and thus disproportionate sensitivity to how much incremental progress was observed at each check-in when evaluating how quickly a target attribute had progressed (Study 8).

Considering the MFE in Light of Related Literatures

Myopic Loss Aversion

We considered monitoring contexts in which attributes change in one direction. Diseases spread. Workers produce more widgets.

Table 15
Actual and Recalled Target Attribute Values (Study 8)

Measure employee	Actual value	*M* (*SD*)	*t*	*p*	*d*
Recalled observed output					
F	4	4.83 (2.16)	6.40	<.001	0.39
R	12	9.88 (3.19)	−11.02	<.001	0.66
Recalled monitoring frequency					
F	12	10.19 (2.38)	−12.59	<.001	0.76
R	4	5.05 (2.37)	7.36	<.001	0.44

Note. The test statistics, significance values, and effect sizes correspond to one-sample *t* tests comparing the recalled attribute values against the actual value for a particular target. F = frequent; R = rare.

Children grow taller. Under a more watchful eye, change seemed to slow. Monitors became too focused on what they observed at the moment (little progress since the last check-in) and neglected that each of these many check-ins produced big changes in the aggregate. But in other contexts, the very direction of change can be up in the air. Even if a stock yields positive returns on the year, there will be certain days on which its value will decline. People are often advised not to track their retirement portfolios on a daily basis, because such monitoring frequency increases the likelihood that people will observe (and be spooked by) negative changes in their investment totals. This can produce myopic loss aversion (Blavatskyy & Pogrebna, 2009; Thaler et al., 1997): a reluctance to take (actually beneficial) risks that comes from short-sided reactions to losses that are observable with frequent monitoring (e.g., on days the stock market is down) but would actually be corrected for or even missed in the aggregate (e.g., when looking only at positive annual returns). Both myopic loss aversion and the MFE highlight how frequent monitoring can lead one to miss the forest for the trees. With more vigilance, one will notice more negative deviations and see more periods of relative stagnancy. What people neglect is the longer time horizon during which those losses will be canceled out (making myopic loss aversion unwise) and that those many periods of superficial sluggishness will combine to produce more sizable change.

Accuracy When Tracking Frequency

We found that bias in tracking and recalling information about the targets—(somewhat distorted) recollections of how much output tended to be observed at check-ins and how frequently they were monitored—was not a key mechanism that could explain the MFE. That said, especially in light of research that suggests that frequency tends to be automatically processed (e.g., Flexser & Bower, 1975; Hasher et al., 1987; Howell, 1973; for a review, see Hasher & Zacks, 1984), readers may be surprised that we found systematic errors in these recollections. But being able to track such information automatically and thus efficiently does not mean that it will be tracked perfectly. When one is monitoring the progress of several targets, then even minor errors of source confusion will lead the targets to be recalled as more similar than they would have been otherwise. In fact, when for each recalled attribute one sums the two recalled values, we actually see stronger evidence of recall accuracy. For example, participants on average recalled monitoring the workers 15.24 times, which shows little systematic bias from the 16 total check-ins. Participants instead simply recalled the targets as being slightly more similar on these attributes (observed output, monitoring frequency) than they actually were. But because this source confusion was equally pronounced for each attribute, such biases essentially canceled each other out and thus did not—in and of themselves—lead to the MFE.

Ratio Bias

Instead, the MFE was localized to a failure to integrate the recalled attribute values in a normatively defensible way. More specifically, judgments of productivity were independently predicted by the recalled amount of progress that was observed at the check-ins, but not by how often participants remembered completing such check-ins. Had participants synthesized this information accurately, each should have been a positive predictor when controlling for the other. In this way, the MFE has similarities to research on the ratio bias and denominator neglect (e.g., Reyna & Brainerd, 2008), in which judgments are sensitive to focal top-line attribute indices but fail to incorporate the less salient attributes by which these salient markers must be divided to understand their meaning. Such denominator neglect explains why Kokis et al. (2002) found that on 43% of trials participants chose to draw from an urn that had eight or nine winning balls out of 100 instead of an objectively superior one that had just one winning ball out of 10. Denominator neglect reduces with age and intellectual ability (Ferreira et al., 2016; Toplak et al., 2014) but is still prevalent among adults (Acredolo et al., 1989; Reyna & Brainerd, 1993).

We think the MFE is compatible with but distinct from demonstrations of ratio bias. After all, ratio bias studies actually present participants with ratios (e.g., nine winners out of 100 balls) and test whether the denominator is relatively neglected. As our normative productivity calculation made clear, progress in our paradigms is perhaps most intuitively captured as a sum (of the amounts observed at each check-in) or a product (i.e., The Amount Observed at a Typical Check-In × The Number of Check-Ins), a representation that is not possible in the classic ratio-bias paradigms. Furthermore, in our paradigms, we held constant the total duration that targets were monitored. Had we instead varied this duration—for example, having participants monitor one target for 12 days and another for 24 days—productivity would need to be represented mentally (even if not by the experimenter) as a quotient (total output/days of work). In that case, the worker monitored for longer might appear more productive, because the implicit numerator (total output) may not be adequately adjusted in light of the implicit[7] denominator (total days worked).

Despite these differences, we think the MFE and the ratio bias highlight complementary ways in which certain attributes are naturally salient and thus exert disproportionate influence on more complex judgments. The representativeness heuristic (Kahneman & Tversky, 1972)—with its emphasis that people judge complex features by leaning on judgments of superficially similar features—hints at a useful umbrella theoretical concept that applies to both the MFE and denominator neglect. Much as the perceived attractiveness of a gamble may be disproportionately guided by the number of opportunities to win (and neglect the total number of balls that shape the likelihood that those winners will be drawn), judgments of (to take one of our paradigms) productivity disproportionately weight observations of worker production (and neglect the time scope over which such observations were collected). In all cases, perceivers are sensitive to the attribute that is most similar to the characteristic being judged (i.e., desirability, productivity) and neglect the background feature that aids in interpreting this focal attribute.

Partitioning

The MFE explores the consequences of dividing a time period over which monitoring will occur into many small windows (when monitoring is frequent) or into just a few longer windows (when monitoring is rare). We were interested in such partitioning as it related to information acquisition. This contrasts with the literature on partition dependence, in which response categories are divided into coarser or more fine-grained response options. Response patterns are biased toward filling those categories equally (Fox & Clemen, 2005). Even professional senior auditors, for example, will see accounting estimates in a certain range as more or less reasonable if they are asked to evaluate many or just a few alternatives in that range (Wolfe et al., 2017). The MFE thus illustrates a qualitatively different role of partitioning.

The MFE, which documents the consequences of dividing monitoring into subperiods, may appear to be at odds with effects of subadditivity and unpacking, whereby breaking down larger categories into subcategories increases the perceived magnitude or frequency of those overarching categories (e.g., Fiedler et al., 2009; Mulford & Dawes, 1999). For example, Tversky and Koehler (1994) found that participants estimated that 58% of American deaths are due to natural causes. When the packed category "natural causes" was unpacked for participants into component parts—"heart disease, cancer, and other natural causes"—the estimate went up by 15 percentage points. But with the MFE, breaking up total progress into more component parts leads the overall progress to seem *smaller*. We suspect the superficial discrepancy is attributable to a difference in the paradigms used to study (and thus the mechanisms that give rise to) these distinct phenomena. In the unpacking literature, unpacking an amorphously large category into component parts helps to clarify and make concrete just how many varied components serve to define the overarching (packed) category. In contrast, in our paradigms, participants were always directly exposed to the full quantity of progress (e.g., the number of widgets produced, the number of infection cases). In that sense, the entire category (e.g., overall productivity) was always fully unpacked; what was varied was merely how many of these unpacked pieces were summed up to characterize the feedback at a single check-in. In this sense, the MFE and unpacking are not antagonistic, but merely share superficial features that at first glance can mask the distinct underlying processes that give rise to them.

Goal Striving

We studied the MFE in contexts in which perceivers track information about social and nonsocial targets. But as several of our rhetorical examples have highlighted, people also engage in self-tracking. This becomes particularly relevant in the context of goal striving, in which a sense that one is making progress toward an end state quickly or slowly has been shown to have motivational consequences. That said, there is inconsistent evidence regarding the directionality of those effects.

On the one hand, when people feel their goal progress is slow, they may experience negative affect that can encourage the discontinuation of goal pursuit (Carver & Scheier, 2001; Clore et al., 2001). To avoid the well-being loss that can come from clinging to seemingly unattainable goals (Wrosch et al, 2003), people may experience an action crisis that encourages them to lower the ambitiousness of their goals (Elmi et al., 2023), disengage from them (Heckhausen et al., 2010), or abandon them altogether (Brandstätter & Schüler, 2013). For example, to the extent that a dieter weighs themselves especially frequently, their sense of slow progress may undermine their self-efficacy and undermine their perseverance (Bagozzi & Dholakia, 1999). But according to motivation intensity theory, people will increase their effort when they feel they are facing difficulties reaching their end state (Brehm, 1975; Brehm & Self, 1989).

One possible resolution is that both of these theories may be correct, but for different people. For example, those who engage in positive self-talk in light of goal setbacks—a quality that tends to grow with age (Lam & Zhou, 2022)—are those who tend to ultimately attain their goals (Hennecke et al., 2019). Whether these same individuals are most resilient in the face of a (self-imposed) psychological setback that can come from frequent monitoring (i.e., the MFE) is an open question for future research.

Monitoring frequency joins other forces—some (like monitoring frequency) cognitive, some motivational—that can affect perceptions of goal progress. For example, Huang et al. (2012) reported that people are motivated to see more goal progress than they have actually achieved when they are early in the goal pursuit process, for such exaggerated perceptions will encourage continued goal commitment. Taking a more cognitive angle, Sharif and Woolley (2020) showed that people are sensitive to how many categories of steps (e.g., completing the first of two sets of exercises) and not to how many components compose each category (e.g., the completed set may contain only a small fraction of the number of exercises that the remaining set contains). Unlike these examples, the MFE most directly implicates the apparent speed with which progress is being achieved. But when—like in our studies—the total time over which such monitoring occurs is held constant, it also follows that monitoring frequency reduces a sense that goal progress has been made.

[7] We refer to these as implicit to communicate that they need not be presented as a numerator and denominator, but their normative application would require them to be used in these roles.

There has been some research on how goals themselves can affect monitoring frequency, and the present findings may shed light on those results. Chang, Webb, Benn, and Reynolds (2017) had participants take part in a financial management simulation with two distinct goals in mind. One was a promotion-focused goal (to reach a certain high monetary threshold), while the other was a prevention-focused goal (not to fall below a low monetary threshold). When participants felt the promotion-focused goal was particularly important, a sense that they were doing poorly reduced monitoring frequency (of their own standing). When instead participants thought the prevention-focused goal was important, poor performance increased monitoring. By the MFE, poor performers who were focused on promotion (and thus were infrequent monitors) may have come to see themselves as getting out of their slump. Poor performers who were focused on prevention (and thus were frequent monitors) may have continued to feel mired in it. This may have produced (unjustified) optimism or pessimism about their prospects. Ultimately, how the confidence-boosting (vs. confidence-reducing) effect of monitoring frequency balances against the information that frequent monitoring provides to self-regulatory efforts likely depends on the performance domain. If this trade-off were understood in a particular performance context, people could perhaps be nudged to adopt the *same* goal but in promotion- or prevention-focused terms, based on which sort of monitoring approach would prove most adaptive.

Theoretical Contribution

The Counterintuitive Nature of the MFE

The MFE joins a set of psychological phenomena that people readily display, but less readily intuit. In general, such counterintuitive findings fall in two categories. Some findings—like affective forecasting errors (Wilson & Gilbert, 2005) and perceptions of intergroup polarization (Westfall et al., 2015)—are those for which people understand the general direction of an effect (e.g., Christmastime is a joyous season for its celebrants …), but misperceive its magnitude (… but not as joyous as people think it will be; Buehler & McFarland, 2001). For other phenomena—like the satisfaction that comes from forging connections with random strangers (Epley & Schroeder, 2014) or the greater propensity to return lost property that is of greater monetary value (Cohn et al., 2019)—people misintuit even the direction of the effect. The present article offered evidence that the MFE falls into this latter category. This not only helps to further delineate the classes of psychological phenomena into which people have especially poor insight, but it identifies a likely source of impression mismanagement. Those who hope to 1 day achieve the spotlight may do best by working in the shadows. Of course, once the fruits of one's labor are achieved, social attention should be beneficial, assuming those fruits are indeed sweet.

A Novel Way Beliefs and Expectations Guide Performance Evaluations

The MFE identifies a previously unidentified route by which prior expectations and beliefs about targets may change how those targets' actions are evaluated. Although Study 2a identified a handful of target features that change how frequently targets are monitored, recall that one of those related to whether people started with positive or negative expectations about the target. There are many factors—including social category cues like ethnicity (Papageorge et al., 2020), socioeconomic status (Figlio, 2005), and gender (Carlana, 2019)—that are known to guide expectations about a target's abilities and performance. Sometimes such expectancies can produce actual changes in performance. When teachers have higher (vs. lower) expectations, students seem to perform better (vs. worse; e.g., Carlana, 2019; Figlio, 2005; Gentrup et al., 2020; Hill & Jones, 2021; Papageorge et al., 2020; for reviews, see Jussim & Harber, 2005; S. Wang et al., 2018). This causal influence can emerge when expectations shape, for example, the frequency or nature of the feedback that teachers provide to a student for whom they do or do not have high hopes (Gentrup et al., 2020). In this way, changes to actual performance are a first way in which expectations can change performance evaluations.

Other routes, like the one identified by the MFE, do not require an actual change in performance. For example, beliefs and expectations can encourage people to seek out information that has the potential to confirm those priors (e.g., Rajsic et al., 2015). Due to this confirmation bias, perceivers reinforce their initial beliefs. By a separate process, beliefs and expectations can lead people to evaluate or interpret the *same* evidence in a way that reinforces their prior beliefs. Such effects do not simply reflect the stickiness of a prior, for these expectancies color judgments only or especially when they are set *before* perceivers actually sample the evidence. For example, Lee et al. (2006) found that people liked the taste of beer better when a few drops of a secret ingredient had been added. However, if participants knew before tasting that the addition was balsamic vinegar, then this disgusting-sounding supplement contaminated the subsequent tasting experience. But crucially, learning this feature afterward did not spoil the memory of the already completed trial. Similarly, Critcher and Dunning (2009) showed more directly how a priori expectations on a performance task can color one's bottom-up experience while completing the task (e.g., fueling a sense that one was struggling less on each question or solving the problems more quickly), thereby explaining why prior self-views can color performance evaluations (Ehrlinger & Dunning, 2003).

Consider how the MFE reflects a qualitatively distinct way in which beliefs or expectations can color perceptions of a target. Previously identified mechanisms highlight how these priors can bring about actual changes in a target, can modify what information one attends to, or can change how such ambiguous information is interpreted, all serving to confirm one's existing priors. The MFE instead identifies a more indirect route, one by which beliefs or expectations can change the pace at which the same information is gathered, which in turn affects perceptions. More generally, to the extent that different beliefs and expectations can influence the frequency of monitoring a target, then the MFE establishes why these factors can serve as causal precursors to perceptions of targets' rate of change and thus propensity for progress.

Broader Implications

The MFE may offer insight into how factors beyond beliefs and expectations color evaluations of progress. For example, motivation and personal relevance could also influence monitoring frequency, with predictable consequences for evaluations. We observed one example of this in Study 2a, in which participants indicated a preference for more frequent monitoring of more important projects. The perverse effect whereby task importance encourages more frequent monitoring, which (by the MFE) should slow perceptions of progress, presumably extends to self-monitoring as well. This suggests a perverse trap. The dieter who is eager to lose weight, the person who is desperate to gain a social media following, and the academic who aspires to be influential in the field may find themselves frequently checking how many kilos they have lost, how many followers they have acquired, and how many new citations they have amassed. Such eagerness may ultimately beget a sense of failure and inadequacy. We ourselves will be frequently checking to see whether our own work inspires many direct tests of these ideas.

Constraints on Generality

We observed consistent support for the MFE across three participant populations: undergraduates at an American university, undergraduates at a Portuguese university, and a more general population of Americans recruited from online platforms. We do not have reason to believe that the MFE would be limited to these populations. In all of our studies, change occurred in a single direction. We thus urge caution in attempting to extrapolate our findings to contexts in which change is not monotonic. The MFE emerged in a variety of contexts that used both social and nonsocial targets. That said, in many of our paradigms, participants likely had little-to-no prior experience tracking changes in the relevant targets. One obvious exception is Study 6, in which participants tracked changes in height, and it is notable that the MFE remained robust. Still, we do urge caution in extrapolating our findings to perceivers who are more expert in the tracking domain. For example, a professional epidemiologist may be better at tracking the actual rate of growth of a contagious disease than our participants in Study 3 were. That said, to the extent that public actions and interventions are informed by both expert and novice trackers (who may, for example, give support to government-supported initiatives to address environmental, public health, and other problems), we think the results still have practical importance.

References

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development*, *60*(4), 933–945. https://doi.org/10.2307/1131034

Alves, H., & Mata, A. (2019). The redundancy in cumulative information and how it biases impressions. *Journal of Personality and Social Psychology*, *117*(6), 1035–1060. https://doi.org/10.1037/pspa0000169

Alves, H., Vogel, T., Grüning, D., & Mata, A. (2023). Why leading is (almost) as important as winning. *Cognition*, *230*, Article 105282. https://doi.org/10.1016/j.cognition.2022.105282

Atlas, S. A., & Bartels, D. M. (2018). Periodic pricing and perceived contract benefits. *Journal of Consumer Research*, *45*(2), 350–364. https://doi.org/10.1093/jcr/ucx127

Bagozzi, R. P., & Dholakia, U. (1999). Goal setting and goal striving in consumer behavior. *Journal of Marketing*, *63*(4, Suppl. 1), 19–32. https://doi.org/10.1177/00222429990634s104

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

Blavatskyy, P., & Pogrebna, G. (2009). Myopic loss aversion revisited. *Economics Letters*, *104*(1), 43–45. https://doi.org/10.1016/j.econlet.2009.03.023

Brandstätter, V., & Schüler, J. (2013). Action crisis and cost–benefit thinking: A cognitive analysis of a goal-disengagement phase. *Journal of Experimental Social Psychology*, *49*(3), 543–553. https://doi.org/10.1016/j.jesp.2012.10.004

Brehm, J. W. (1975). *A theory of motivational suppression* (Research proposal submitted to the National Science Foundation). Department of Psychology, University of Kansas.

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Buehler, R., & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin*, *27*(11), 1480–1493. https://doi.org/10.1177/01461672012711009

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, *134*(3), 1163–1224. https://doi.org/10.1093/qje/qjz008

Carver, C. S., & Scheier, M. F. (2001). Optimism, pessimism, and self-regulation. In E. C. Chang (Ed.), *Optimism & pessimism: Implications for theory, research, and practice* (pp. 31–51). American Psychological Association. https://doi.org/10.1037/10385-002

Chang, B. P. I., Webb, T. L., & Benn, Y. (2017). Why do people act like the proverbial ostrich? Investigating the reasons that people provide for not monitoring their goal progress. *Frontiers in Psychology*, *8*, Article 152. https://doi.org/10.3389/fpsyg.2017.00152

Chang, B. P. I., Webb, T. L., Benn, Y., & Reynolds, J. P. (2017). Monitoring personal finances: Evidence that goal progress and regulatory focus influence when people check their balance. *Journal of Economic Psychology*, *62*, 33–49. https://doi.org/10.1016/j.joep.2017.05.003

Chang, B. P. I., Webb, T. L., Benn, Y., & Stride, C. B. (2017). Which factors are associated with monitoring goal progress? *Frontiers in Psychology*, *8*, Article 434. https://doi.org/10.3389/fpsyg.2017.00434

Clore, G. L., Gasper, K., & Garvin, E. (2001). Affect as information. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of affect and social cognition* (pp. 121–144). Sage Publications. https://doi.org/10.4135/9781412956253.n8

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, *365*(6448), 70–73. https://doi.org/10.1126/science.aau8712

Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, *97*(6), 931–945. https://doi.org/10.1037/a0017452

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*(1), 5–17. https://doi.org/10.1037/0022-3514.84.1.5

Elmi, C. M., Napolitano, C. M., & Freund, A. M. (2023). Shift happens: Changes in goal orientation during goal pursuit. *European Psychologist*, *28*(1), 53–63. https://doi.org/10.1027/1016-9040/a000496

Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology: General*, *143*(5), 1980–1999. https://doi.org/10.1037/a0037323

Ferreira, M. B., Mata, A., Donkin, C., Sherman, S. J., & Ihmels, M. (2016). Analytic and heuristic processes in the detection and resolution of conflict. *Memory & Cognition*, *44*(7), 1050–1063. https://doi.org/10.3758/s13421-016-0618-7

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. *Psychology of Learning and Motivation*, *57*, 1–55. https://doi.org/10.1016/B978-0-12-394293-7.00001-7

Fiedler, K., Hofferbert, J., & Wöllert, F. (2018). Metacognitive myopia in hidden-profile tasks: The failure to control for repetition biases.

Frontiers in Psychology, *9*, Article 903. https://doi.org/10.3389/fpsyg.2018.00903

Fiedler, K., Unkelbach, C., & Freytag, P. (2009). On splitting and merging categories: A regression account of subadditivity. *Memory & Cognition*, *37*(4), 383–393. https://doi.org/10.3758/MC.37.4.383

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom—A cognitive-environmental approach. *Organizational Behavior and Human Decision Processes*, *88*(1), 527–561. https://doi.org/10.1006/obhd.2001.2981

Figlio, D. N. (2005). *Names, expectations and the black–white test score gap*. National Bureau of Economic Research. https://doi.org/10.3386/w11195

Flexser, A. J., & Bower, G. H. (1975). Further evidence regarding instructional effects on frequency judgments. *Bulletin of the Psychonomic Society*, *6*(3), 321–324. https://doi.org/10.3758/BF03336675

Fox, C. R., & Clemen, R. T. (2005). Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science*, *51*(9), 1417–1432. https://doi.org/10.1287/mnsc.1050.0409

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55. https://doi.org/10.1037/0022-3514.65.1.45

Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, *66*, Article 101296. https://doi.org/10.1016/j.learninstruc.2019.101296

Gourville, J. T. (1998). Pennies-a-day: The effect of temporal reframing on transaction evaluation. *Journal of Consumer Research*, *24*(4), 395–408. https://doi.org/10.1086/209517

Grüning, D. J., Alves, H., Mata, A., & Fiedler, K. (2023). Reversing the cumulative redundancy bias to demonstrate metacognitive flexibility in cue utilization. *Journal of Experimental Social Psychology*, *107*, Article 104471. https://doi.org/10.1016/j.jesp.2023.104471

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388. https://doi.org/10.1037/0003-066X.39.12.1372

Hasher, L., Zacks, R. T., Rose, K. C., & Sanft, H. (1987). Truly incidental encoding of frequency information. *The American Journal of Psychology*, *100*(1), 69–91. https://doi.org/10.2307/1422643

Heckhausen, J., Wrosch, C., & Schulz, R. (2010). A motivational theory of life-span development. *Psychological Review*, *117*(1), 32–60. https://doi.org/10.1037/a0017668

Hennecke, M., Czikmantori, T., & Brandstätter, V. (2019). Doing despite disliking: Self-regulatory strategies in everyday aversive activities. *European Journal of Personality*, *33*(1), 104–128. https://doi.org/10.1002/per.2182

Hill, A. J., & Jones, D. B. (2021). Self-fulfilling prophecies in the classroom. *Journal of Human Capital*, *15*(3), 400–431. https://doi.org/10.1086/715204

Howell, W. C. (1973). Storage of events and event frequencies: A comparison of two paradigms in memory. *Journal of Experimental Psychology*, *98*(2), 260–263. https://doi.org/10.1037/h0034380

Hsee, C. K., Yang, Y., & Li, X. (2019). Relevance insensitivity: A new look at some old biases. *Organizational Behavior and Human Decision Processes*, *153*, 13–26. https://doi.org/10.1016/j.obhdp.2019.05.002

Hsee, C. K., Zhang, J., Lu, Z. Y., & Xu, F. (2013). Unit asking: A method to boost donations and beyond. *Psychological Science*, *24*(9), 1801–1808. https://doi.org/10.1177/0956797613482947

Huang, S. C., Zhang, Y., & Broniarczyk, S. M. (2012). So near and yet so far: The mental representation of goal progress. *Journal of Personality and Social Psychology*, *103*(2), 225–241. https://doi.org/10.1037/a0028443

Jenkins, J., & Terjeson, K. J. (2011). Monitoring reading growth: Goal setting, measurement frequency, and methods of evaluation. *Learning Disabilities Research & Practice*, *26*(1), 28–35. https://doi.org/10.1111/j.1540-5826.2010.00322.x

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (pp. 693–708). Cambridge University Press/Russell Sage Foundation. https://doi.org/10.1017/CBO9780511803475.039

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228. https://doi.org/10.1037/0033-295X.94.2.211

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*(1), 26–52. https://doi.org/10.1016/S0022-0965(02)00121-2

Lam, K. K. L., & Zhou, M. (2022). An examination of the conceptual structure of long-term goal striving. *Personality and Social Psychology Bulletin*, *48*(4), 550–565. https://doi.org/10.1177/01461672211016190

Lee, L., Frederick, S., & Ariely, D. (2006). Try it, you'll like it: The influence of expectation, consumption, and revelation on preferences for beer. *Psychological Science*, *17*(12), 1054–1058. https://doi.org/10.1111/j.1467-9280.2006.01829.x

Li, X., & Hsee, C. K. (2021). The psychology of marginal utility. *Journal of Consumer Research*, *48*(1), 169–188. https://doi.org/10.1093/jcr/ucaa064

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Mulford, M., & Dawes, R. M. (1999). Subadditivity in memory for personal events. *Psychological Science*, *10*(1), 47–51. https://doi.org/10.1111/1467-9280.00105

Papageorge, N. W., Gershenson, S., & Kang, K. M. (2020). Teacher expectations matter. *The Review of Economics and Statistics*, *102*(2), 234–251. https://doi.org/10.1162/rest_a_00838

Paz, I. G. (2021, June 4). Florida will no longer publish daily coronavirus reports. *The New York Times*. https://www.nytimes.com/2021/06/04/world/florida-covid-19-cases-vaccine.html

Rajsic, J., Wilson, D. E., & Pratt, J. (2015). Confirmation bias in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(5), 1353–1364. https://doi.org/10.1037/xhp0000090

Read, D., Loewenstein, G., Rabin, M., Keren, G., & Laibson, D. (1999). Choice bracketing. In B. Fischoff & C. F. Manski (Eds.), *Elicitation of preferences* (pp. 171–202). Springer. https://doi.org/10.1007/978-94-017-1406-8_7

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3–8. https://doi.org/10.1016/0304-3959(96)02994-6

Reyna, V. F., & Brainerd, C. J. (1993). Fuzzy memory and mathematics in the classroom. In G. M. Davies & R. H. Logie (Eds.), *Memory in everyday life* (pp. 91–119). North Holland Press. https://doi.org/10.1016/S0166-4115(08)61096-1

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*(1), 89–107. https://doi.org/10.1016/j.lindif.2007.03.011

Ritchie, H., Rosado, P., & Roser, M. (2023). *CO_2 and greenhouse gas emissions* [Data set]. Our World in Data. https://ourworldindata.org/co2-and-greenhouse-gas-emissions

Sharif, M. A., & Woolley, K. (2020). The effect of categorization on goal progress perceptions and motivation. *Journal of Consumer Research*, *47*(4), 608–630. https://doi.org/10.1093/jcr/ucaa022

Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics*, *112*(2), 647–661. https://doi.org/10.1162/003355397555226

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology*, *50*(4), 1037–1048. https://doi.org/10.1037/a0034910

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*(4), 547–567. https://doi.org/10.1037/0033-295X.101.4.547

Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes*, *103*(1), 37–52. https://doi.org/10.1016/j.obhdp.2006.12.002

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*(3), 169–185. https://doi.org/10.1002/bdm.3960050303

Vedejová, D., & Čavojová, V. (2021). Confirmation bias in information search, interpretation, and memory recall: Evidence from reasoning about four controversial topics. *Thinking & Reasoning*, *28*(1), 1–28. https://doi.org/10.1080/13546783.2021.1891967

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Wang, Y., Baum, S. M., & Critcher, C. R. (2023). Needing everything (or just one thing) to go right: Myopic preferences for consolidating or spreading risks. *Journal of Personality and Social Psychology*, *125*(4), 730–751. https://doi.org/10.1037/pspa0000343

Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology*, *92*(5), 821–833. https://doi.org/10.1037/0022-3514.92.5.821

Webb, T. L., Chang, B. P. I., & Benn, Y. (2013). "The ostrich problem": Motivated avoidance or rejection of information about goal progress. *Social and Personality Psychology Compass*, *7*(11), 794–807. https://doi.org/10.1111/spc3.12071

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, *10*(2), 145–158. https://doi.org/10.1177/1745691615569849

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131–134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wolfe, C. J., Fitzgerald, B. C., & Newton, N. J. (2017). The effect of partition dependence on assessing accounting estimates. *Auditing: A Journal of Practice & Theory*, *36*(3), 185–197. https://doi.org/10.2308/ajpt-51666

Wrosch, C., Scheier, M. F., Carver, C. S., & Schulz, R. (2003). The importance of goal disengagement in adaptive self-regulation: When giving up is beneficial. *Self and Identity*, *2*(1), 1–20. https://doi.org/10.1080/15298860309021

Yang, Y., Hsee, C. K., & Li, X. (2021). Prediction biases: An integrative review. *Current Directions in Psychological Science*, *30*(3), 195–201. https://doi.org/10.1177/0963721421990341

Received April 16, 2023
Revision received December 18, 2024
Accepted December 21, 2024 ■