*Psychological Science*, in press

Illusory Feelings, Elusive Habits: Explanations of Behavior Overlook Habits

Asaf Mazar and Wendy Wood

University of Southern California

**Abstract**

Habits underlie much of human behavior. However, people may prefer agentic explanations that overlook habits in favor of inner states such as mood. We tested this misattribution hypothesis in an online experiment of helping behavior as well as an ecological momentary assessment study of college students’ everyday coffee drinking. Both studies revealed a substantial gap between attributed and actual influences on behavior: Habit strength outperformed or matched inner states in predicting behavior, whereas participants’ attributions for their behavior emphasized inner states. Participants continued to overlook habits even when incentivized for accuracy, as well as when making attributions for other people’s behavior. We discuss how this attribution pattern could adversely influence self-regulation.

*Keywords:* habit, attribution, automaticity, lay theories, cognitive bias

**Statement of Relevance**

Habits are ubiquitous in daily life, but people often prefer to explain their behavior using inner states. Thus, they may overlook the influence of habit. In two studies, we tested this bias to overlook habits in favor of inner states by comparing participants’ perceptions with the actual predictors of two behaviors: helping others in a lab experiment and drinking coffee in daily life. Participants attributed these behaviors more to mood and fatigue than to habit, whereas habit had an equal or stronger influence on actual behavior. With these two behaviors, we demonstrated the bias in a lab study manipulating habit and mood as well as with habits in daily life. To the extent that people overlook habit in this way, they will be ill-equipped to effectively self-regulate habitual behavior.

**Open Practices Statement**

Both studies were pre-registered. All pre-registration plans, materials, de-identified data, and analytic code are hosted on the Open Science Framework at:<https://osf.io/5xfsm/?view_only=ba7cacb9c9e047c1ad60fdd137f6d84a>.

**Illusory Feelings, Elusive Habits: Explanations of Behavior Overlook Habits**

Habits are key to successful functioning in our day-to-day lives. By automating repeated behaviors, habits enable people to consistently get enough sleep, stay fit, eat a healthy diet, and study (Galla & Duckworth, 2015). Furthermore, engaging in routine, habitual behaviors is associated with a greater sense of security and even meaning in life (Avni-Babad, 2011; Heintzlman & King, 2019). Everyday habits also enable multitasking: In an experience sampling study, 43% of daily behaviors were performed habitually in that they were repeated frequently in the same location, typically while participants were thinking about something other than what they were doing (Wood et al., 2002).

Habits are mental associations between contexts and responses that develop as people repeat rewarded responses in a given context (Knowlton & Diedrichsen, 2018). Once habits have formed, context cues automatically activate the repeated response in mind (Mazar & Wood, 2018). Habit associations are separate from the behavior they produce: For example, when a entering a car, the habitual response of wearing a seatbelt may be activated in mind regardless of whether one overtly acts on it by buckling up.

Given the many opportunities people have in daily life to observe their own repeated actions, one might expect lay theories to accurately account for habits. In support, participants in one study read about an office worker who locked in a colleague by turning the office door knob counterclockwise (Gershman et al., 2016). Participants demonstrated attribution to habit by placing less blame on the office worker if their home doors also opened counterclockwise, implying that the office worker acted out of habit. Also relevant, the actor-observer effect posits that people readily attribute their own behavior (but not others’ behavior) to environmental influences (Jones & Nisbett, 1971). Because habits are activated by cues in the environment, people may consequently ascribe their own behavior to habits.

Nevertheless, there is reason to believe that people overlook habit when accounting for their own repeated, habitual behaviors. People overvalue introspective thoughts, feelings, and emotions in self-judgments (Pronin, 2009); and they interpret actions as intentional by default (Rosset, 2008). Illustrating overattribution to inner states, smokers in one study reported that their smoking was triggered by negative affect, even though in-the-moment affect assessments revealed little association between negative affect and subsequent smoking (Shiffman et al., 1997). In another study, self-described emotional eaters demonstrated little daily association between negative affect and snacking (Adriaanse et al., 2011). In addition, participants with stronger habits reported greater certainty in their behavioral intentions, even though these intentions did not predict their future behavior (Ji & Wood, 2007). Taken together, these findings suggest that people may exaggerate the effect of inner states on behavior while discounting the role of habit.

In sum, the present article tests a potential bias to overlook the influence of habits on behavior. Such a bias would be important to document given the many downstream influences of lay theories about behavior (e.g. McFerran & Mukhopadhyay, 2013). For example, such a bias may cause people to ineffectively self-regulate by putting too much weight on regulating inner states such as mood and too little weight on using self-regulation strategies that may better control habits such as reducing triggering cues (e.g. Duckworth et al., 2016).

**The Present Research**

In two studies, we measure the effects of habits and inner states on a behavior and assess participants’ attributions for that behavior. Our first study was an experiment that orthogonally manipulated habit strength and mood to assess their effects on helping behavior. Our second study used ecological momentary assessment to track coffee drinking over a typical week. In both studies, miscalibration of the actual and perceived effects of habit would emerge if participants placed less value on habit than inner states in their attributions than would be accurate, given the actual effects of habit and inner states on behavior.

**Study 1**

Participants first recalled a happy, sad, or neutral event and then completed a simple, supposedly unrelated task that trained either a strong or weak habit to press one of two computer keys. Immediately after the task, participants pressed one of the keys to indicate if they were willing to donate a small amount of time to help the researchers. The habit manipulation should lead strong habit participants to select the response that maps onto the extensively-trained key. Specifically, our hypotheses compared (a) the actual effects of the mood and habit manipulations on participants’ decision to help or not with (b) participants’ attributions for their helping.

Orthogonally manipulating habits and inner states does not imply that habits and inner states are always uncorrelated. Indeed, in daily life, habits often align with moods, goals, and feelings. However, once formed, habitual responses are triggered directly by context cues, with inner states providing limited input (Wood et al., 2021).

**Method**

***Power Analyses***

Power analyses for logistic regression were conducted using the WebPower package in R [(Zhang & Yuan, 2018)](https://www.zotero.org/google-docs/?jDs8fS). Results suggested that 787 participants were required to achieve 80% power for detecting a difference of 10% in helping behavior between conditions (namely, 45% help in one condition vs. 55% in another). Because this study used a novel habit-formation task, this expected difference was chosen because it represents a plausible moderate effect size. Pre-registration plans including exclusion criteria, materials, data files, and analytic code for both studies can be accessed at: <https://osf.io/5xfsm/?view_only=ba7cacb9c9e047c1ad60fdd137f6d84a>

***Participants***

We recruited 809 online participants via Prolific (388 male, 392 female, 15 other, 14 declined to answer; *M*age= 35.17, *SD*age= 13.32, range = 18-78). An additional 115 participants were excluded who did not pass the comprehension check, and an additional 91 participants were excluded due to extreme PANAS scores (+/-2 SD from their mood condition’s mean; 76 participants) or reaction times (median reaction time > 500 ms; 15 participants). Including all participants in the analyses did not notably alter the results (see Table S2 in supplementary materials).

***Procedure***

 After providing informed consent, participants completed an autobiographical emotional memory task [(Mills & D’Mello, 2014)](https://www.zotero.org/google-docs/?NdX6zX) in which they recalled either a happy, sad, or neutral memory. The description of each emotion condition is given in the bolded text in parentheses below:

Recall (**an event in your life that made you happy/ an event in your life that made you sad/ the last time that you brushed your teeth**). Take some time to really experience the event and the feelings associated with it. When you are ready, describe the event below in your own words. You may use between 5-40 words.

Participants had a minimum of 30 seconds to write about the scenario in a text box, after which they could proceed with the experiment. As a manipulation check, participants then completed items from the Positive And Negative Affect Schedule [(Thompson, 2007](https://www.zotero.org/google-docs/?Gy0PEL)).

For the habit-formation task, participants completed 40 trials in which the letter “m” or “z” appeared on the screen, and participants were required to respond by pressing the corresponding key. In the *strong habit condition,* participants practiced either the left (“z”) or the right (“m”) response on 36 out of 40 trials (90%) and the alternate response on the remaining 4 trials (10%). The specific response (z/m) was counterbalanced across participants. In the *weak habit condition,* participants practiced each response equally (20 trials on each side). Participants were instructed to be as fast and accurate as possible. This task follows habit formation procedures in prior research (e.g., Hardwick et al., 2018).

Immediately following the task, a screen displayed the helping request: “Are you willing to complete 40 additional trials (~5 minutes) as a favor to us (without additional compensation),” with “Yes” and “No” responses mapped onto the same keys used in the habit-formation task (“z” and “m”). The behavioral measure was whether participants agreed or not. For strong habit participants, the “No” response was always mapped onto the more heavily practiced response. For weak habits participants, “Yes” and “No” responses were randomly assigned to each key. The “No” response was over-trained in the strong habit condition because the helping request (donating five minutes of participants’ time) was taxing and might have spurred participants to exert control to decline it. Participants then answered a comprehension check to test whether they understood the help request along with additional measures (see below).

***Measures***

  **Positive and Negative Mood.** Participants responded to the prompt “indicate the extent to which you feel the mood below RIGHT NOW.” On scales ranging from 1 (*very slightly/not at all*) to 9 (*extremely),* participants rated positive emotions (”inspired,” ”determined,” ”attentive,” ”proud,” ”alert,” ”active”) and negative ones (**“**upset,” ”hostile,” “ashamed,” ”nervous,” ”afraid,” ”guilty”) taken from the Positive And Negative Affect Schedule [(Thompson, 2007)](https://www.zotero.org/google-docs/?zFa9YL). Positive and negative item ratings were averaged to create a positive affect score (𝛼 = .82) and a negative affect score (𝛼 = .80).

 **Self-Attribution.** On scales ranging from 0% (*not at all important*) to 50% or more (*extremely important*), participants rated the extent to which their decision to help or not was due to habit, “I responded automatically, without thinking,” and mood, “My mood at the time (I felt good/bad).” The sum of both answers could range from 0-100%. Presentation order of mood and habit was counterbalanced in both the attribution and incentivized measures.

 **Incentivized Other-Attribution.** To minimize judgment biases, participants were incentivized to provide accurate explanations for others’ behavior. The incentive should minimize the effects of conversational norms regarding plausible or socially acceptable explanations for a behavior. Attributions for others’ behavior are additionally informative because they should be relatively unaffected by self-serving biases that could influence self-attributions. Thus, on scales ranging from 0% (*not at all important*) to 50% or more (*extremely important*), participants indicated “How important do you think that the following factors are in determining whether OTHER participants agree or decline to complete additional trials?” Participants then rated habit and mood, as in the self-attribution measure (above). Accurate ratings (within 5% of the study results), earned a chance to win a $10 bonus.

 **Habit Strength: Self-Reported Behavior Automaticity Index** (Gardner et al., 2012; a subset of items taken from the Self-Report Habit Index, Verplanken & Orbell, 2003)**.** On scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), participants rated the extent to which “hitting a key (‘z’ or ‘m’) in the task is something that I…”: (a) “did without thinking,” (b) “did automatically,” (c) “did without having to consciously remember,” and (d) “started doing before I realized.” Ratings were averaged to create a perceived automaticity score (𝛼 = .85).

To our knowledge, this is the first use of this measure with a simple finger-movement task, and experienced automaticity did not differ between the weak (*M* = 4.09) and strong habit conditions (*M* = 4.08), 95% CI [-0.25, 0.23], *t*(478.9) = -0.07, *p* = .947, *d* = -.005, 95% CI [-0.15, 0.14]. This measure tests a downstream consequence of habit formation—perceived automaticity. However, perception of automaticity is not an especially sensitive measure in and of itself and could tap processes other than habit (Hagger et al., 2015; Mazar & Wood, 2018). Given that even weak habit participants reported relatively high levels of experienced automaticity, our simple key-pressing task produced uniformly high experienced automaticity. For this reason, we do not discuss this measure further.

 **Habit Strength: Reaction Time (RT).** As in prior research, the strength of habit associations in key-pressing tasks can be assessed directly through reaction times to respond to the cue (Hardwick et al., 2018). Due to the skewness common in reaction time distributions, median rather than mean reaction times were used in all analyses.

 **Comprehension Check.** This measure was included to ensure that participants understood the help request. Immediately after indicating their response, participants answered: **“**What was the request that you just responded to?” by choosing one of the following options: (a) *To continue for an additional 20 minutes*, (b) *To continue for an additional 5 minute*s, (c) *To recommend the study to a friend*, or (d) *To receive double compensation for my participation*. Answers a, c, and d were coded as incorrect.

**Results**

 Descriptive statistics and correlations among key variables are presented in Table 1. The proportion of participants in each condition who agreed to help is presented in Table 2.

**Table 1**

*Study 1: Means (M), Standard Deviations (SD), and Correlations with 95% Confidence Intervals*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | *M (SD)* | 1 | 2 | 3 | 4 | 5 | 6 |
| 1. Habit attribution | 11.48 (13.90) |   |   |   |   |   |   |
| 2. Mood attribution | 28.54 (16.11) | .04[-.03, .11] |   |   |   |   |   |
| 3. Incentivized. habit other-attribution  | 21.35 (13.63) | .49\*\*[.43, .54] | .12\*\*[.05, .19] |   |   |   |   |
| 4. Incentivized mood other-attribution  | 31.99 (12.86) | .04[-.03, .11] | .56\*\*[.51, .61] | .09\*\*[.02, .16] |   |   |   |
| 5. Positive affect | 5.26 (1.52) | .05[-.02, .12] | .11\*\*[.05, .18] | .04[-.02, .11] | .16\*\*[.09, .23] |   |   |
| 6. Negative affect | 2.13 (1.15) | .13\*\*[.06, .19] | .07[-.00, .13] | .11\*\*[.04, .18] | .02[-.05, .09] | -.20\*\*[-.26, -.13] |   |
| 7. Reaction time | 298.8 (61.47) | .05[-.02, .12] | .08\*[.01, .15] | .01[-.06, .08] | .06[-.01, .13] | .15\*\*[.08, .22] | -.04[-.11, .03] |

*Note.* Attributions range from 0 - 50 with higher scores reflecting greater importance, mood scores range from 1-9 with higher numbers reflecting stronger feelings, and higher reaction times (in milliseconds) reflects slower responding in the habit training task.

\* *p* < .05

\*\* *p* < .01

**Table 2**

*Percent (Raw Number) of Participants Agreeing to Help by Mood and Habit Condition*

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Habit condition |  |
|  |  | Weak | Strong | Total |
| Mood condition | Control | 60% (50/83) | 43% (77/178) | 49% (127/261) |
| Sad | 48% (46/96) | 47% (87/185) | 47% (133/281) |
| Happy | 55% (50/91) | 47% (82/176) | 49% (132/267) |
| Total |  | 54% (146/270) | 46% (246/539) | 48% (392/809) |

*Note.* Higher percentages reflect more participants agreeing to help by working 5 extra minutes.

Regression models were fit to the data using the following predictors: mood condition (dummy-coded, with the control condition as the reference level), habit condition (effects-coded, *-1* = weak habit, *+1* = strong habit), and interactions between mood and habit. This analytic design was used for all Study 1 analyses.

***Manipulation Checks on Mood***

The mood induction successfully invoked positive affect (measured via the PANAS): In the sad memory condition, participants reported less positive affect compared with the control condition, *b* = -0.38, 95% CI [-0.65, -0.11], *β* = -0.25, *p* = .006. In the happy memory condition, participants reported more positive affect compared with the control condition, *b* = 0.42, 95% CI [0.15, 0.69], *β* = 0.27, *p* = .003. Habit condition did not significantly influence positive affect, *b* = -0.08, 95% CI [-0.27, 0.12], *β* = -0.05, *p* = .424. No interactions emerged between habit and mood conditions, both *p*s > .2.

Results for negative affect also indicated the success of the mood induction: Participants in the sad condition reported higher levels of negative affect compared with the control condition, *b* = 0.83, 95% CI [0.63, 1.02], *β* = 0.72, *p* < .001. Participants in the happy condition did not differ on negative affect from participants in the control condition, *b* = 0.03, 95% CI = [-0.17, 0.22], *β* = 0.02, *p* = .786. This is in line with previous work showing the relative independence of the positive and negative PANAS subscales, reflecting an underlying independence of positive and negative affect [(e.g., Thompson, 2007)](https://www.zotero.org/google-docs/?RfevX5). Habit condition did not significantly affect negative affect, *b* < 0.01, 95% CI [-0.14, 0.14], *β* < 0.01, *p* = .954. There were no interactions between habit and mood conditions, both *p*s > .2.

As an additional mood manipulation check, two coders blind to condition and hypotheses coded each open-text response to the mood manipulation on (a) whether it displayed negative affect, and (b) whether it displayed positive affect. Coders’ ratings showed high interrater reliability: The negative affect ratings showed 97% agreement, Cohen’s *k* = .95, and positive affect ratings showed 95% agreement, Cohen’s *k* = .91. Disagreements between coders were resolved by discussion. Suggesting that the mood manipulation was successful, most open-text responses in the negative affect condition showed negative affect (84.34%), compared with few of the responses in the control (12.26%) and positive (3.75%) conditions. Similarly, most responses in the positive affect condition showed positive affect (80.15%) compared with fewer in the control (35.25%) and negative (3.56%) conditions.

***Manipulation Check on Habit***

The reaction time measure revealed that the habit manipulation was successful: In the strong habit condition, participants were significantly faster to respond than in the weak habit condition, *b* = -20.43, 95% CI [-28.10, -12.77], *β* = -0.33, *p* < .001. There was a non-significant trend for slower reaction times for participants in the happy mood condition compared with the control condition, *b* = 9.22, 95% CI [-1.47, 19.90], *β* = 0.15, p = .091. No other effects approached significance, all *p*s > .2.

***Habit and Mood Effects on Helping - Actual***

A logistic regression model tested the actual effects of habit and mood on helping behavior (*Yes / No)*. Habit significantly influenced this decision, such that participants in the strong habit condition (who had just extensively practiced the “no” response) were less likely to agree to help, *OR* = 0.84, 95% CI[0.73, 0.98], *p* = .024. Mood condition did not significantly determine helping, either for the sad condition, *OR* = 0.94, 95% CI [0.67, 1.32], *p* = .721, or the happy condition, *OR* = 1.02, 95% CI[0.73, 1.44], *p* = .892. Thus, habit influenced behavior, whereas mood did not.[[1]](#footnote-1)

***Habit and Mood Effects on Helping - Attributed***

To test the perceived effects of habit and mood, a dependent-samples *t-*test assessed the within-person difference in attributions to mood compared with habit. Participants strongly attributed their behavior to mood over habit, *Mdiff* = 17.10, 95% CI [15.66, 18.54], *t*(800) = 23.27, *p* < .001, *d* = 1.14, 95% CI [1.01, 1.26]. As anticipated, a strong albeit somewhat smaller bias in favor of mood remained when participants were incentivized to give accurate attributions for others’ behavior, *Mdiff* = 10.61, 95% CI [9.37, 11.85], *p* < .001, *d* = 0.80, 95% CI [0.69, 0.91]. Thus, as anticipated, participants’ attributions favored mood over habit more than would be expected given the actual effects of each on behavior. It should be noted that despite the strong favoring of mood, habits were judged a plausible explanation, especially for others’ behavior, with habits receiving an importance rating of 21% (maximum 50%), compared with 32% for mood.

***Exploratory Analyses: Intensity of Experience and Attributions***

 We explored whether participants with stronger moods and stronger habits were more likely to make attributions to mood and habit, respectively. In general, participants with stronger internal states were only slightly more likely to attribute their behavior to these states. That is, attribution to mood was weakly correlated with positive PANAS scores, *r*(801) = .11, and marginally with negative scores, *r*(801) = .07 (see Table 1). In addition, attribution to mood was weakly correlated with reaction time on the habit formation task, so that participants with slower RTs gave stronger mood attributions, *r*(801) = .08. Attributions to habit only showed a slight positive correlation with negative affect, *r*(800)= .13.

**Discussion**

 This first experiment provided causal evidence that people’s explanations for their behavior favor inner states over habits even when that behavior is driven by habit. We manipulated habit strength via amount of practice at a key-press task and manipulated mood through a memory recall task. Participants then indicated their willingness to help by pressing a highly practiced or less practiced key.

In our test of actual influences on helping, habit strength determined helping but current mood did not. Specifically, participants in the strong habit condition, who had earlier extensively practiced the “no” response key, were more likely to decline a helping request compared with participants in the weak habit condition, who practiced the “yes” and “no” response keys equally. Thus, participants’ decisions continued to be influenced by their prior key-pressing habit. In contrast, participants induced to feel sad or happy helped at comparable rates as participants in a control condition with no mood manipulation. Note that our hypotheses were not about these behavioral effects per se, but instead concerned the difference between actual and perceived effects of habit and mood on behavior.

When explaining their behavior, participants attributed their helping more to current mood than to habit. Thus, attributions were misaligned with the actual determinants of behavior by underestimating habit effects and overemphasizing mood. Our design provided a compelling test of this hypothesis given that participants in the strong habit condition should be aware of their recent, extensive practice at pressing a particular computer key. However, when incentivized to make accurate attributions about others’ behavior, participants still revealed a substantial attribution gap favoring mood over habit. Thus, it does not seem that the attribution pattern was due to artifacts of social desirability or conversational norms that favor mood explanations. The incentivized measure also suggested that attributions to habit were meaningful: Despite the substantial attribution gap favoring mood over habit, participants considered habit to be a plausible determinant of others’ behavior.

**Study 2**

Study 1 tested explanations for a single behavior immediately following an emotionally evocative experience that might make mood especially salient and thus likely to influence attributions. Study 2 instead investigated the attribution bias with a mundane repeated behavior—coffee drinking—recorded over the course of a typical week. Specifically, coffee drinking was assessed in response to an inner state, fatigue, and to habit strength, which are common reasons for coffee drinking (see pilot below).

Given that some of our participants drank coffee very often, we anticipated that habit would strongly influence actual coffee drinking even more than fatigue. However, as in the first study, we anticipated that participants’ attributions would emphasize fatigue as much as or more than habit. Thus, these two hypotheses together concern the correspondence between the actual and perceived determinants of behavior. In an additional test of our model, we anticipated that fatigue attributions would be unrelated to within-person associations between fatigue and coffee drinking. In other words, people’s beliefs about fatigue determining their coffee drinking will be unrelated to its actual role in driving their individual consumption.

**Method**

***Pilot***

To assess lay beliefs about the causes of coffee drinking, 40 college students (22 male, 16 female, 2 genderqueer or other) rated six causes of coffee drinking on scales ranging from 1 (*not at all important*) to 5 (*extremely important*), including: fatigue (“tiredness or low energy”), habit (“habit or behavior routines”), thirst, taste, social motives (“spending time with friends”) and coffee after a meal. Fatigue was rated as most important (*M* = 4.05, *SD* = 0.96), followed by taste (*M* = 3.58, *SD* = 1.03), habit (*M* = 3.50, *SD* = 1.22), social motives (*M* = 3.17, *SD* = 1.08), having coffee after a meal (*M* = 2.12, *SD* = 1.22), and thirst (*M* = 1.70, *SD* = 0.91). A paired-samples *t*-test comparing fatigue and habit attributions (within participants) revealed that participants attributed coffee drinking to fatigue significantly more than to habit, *M*diff = 0.55, 95% CI [0.14, 0.96], *t*(39) = 2.72, *p* = .01, *d* = 0.50, 95% CI [0.11, 0.88].

***Design***

To capture experiences and explanations as they naturally unfold in daily life, we used a combination of surveys, daily morning reports, and ecological momentary assessment (EMA). Participants first completed intake surveys, including measures of habit strength and attributions for their own coffee drinking. Then, over the course of a week, they reported every two hours on their fatigue and coffee drinking. They also completed a brief survey every morning immediately after waking up.

Our analysis predicted coffee drinking at one prompt from fatigue experienced at the prior prompt. This lagged design minimized any self-report bias that might emerge from concurrent associations between fatigue and coffee drinking (i.e., “I’m drinking coffee therefore I must be tired”). After the study week, participants completed a final survey. Finally, participants completed a follow-up survey once data collection for the study ended.

**Power Analyses.** Ecological Momentary Assessment designs such as the present one produce thousands of prompts (level-1 sample size), which tend to produce very high power for within-person effects. Because most of our research questions could be probed within-person (e.g. using our novel context-specific habit measure), we aimed for a final sample size of 120, which is in line with typical sample sizes in ecological momentary assessment studies (compare with a mean sample size of 99 in a recent systematic review; Wen et al., 2017).

To estimate observed power for our multilevel logistic regression, we simulated a dataset with log-odds regression coefficients of 0.3 and 0.2 (corresponding to our odds-ratios of 1.35 and 1.22) for a level 2 and level 1 variable, respectively. Simulated sampling from this dataset 1,000 times revealed that 50 participants were sufficient to achieve 90% power for our between-subjects variable (habit strength) and 99.5% power for our within-person variable (fatigue).

***Participants***

Participants were a convenience sample of 112 U.S. undergraduate students who received either course credit or monetary compensation (85 female, 27 male, *M*age = 20.85, *SD*age = 2.85, range = 18-33). The (self-reported) selection criteria were: (a) speaking English fluently, (b) owning a smartphone, (c) being 18 or older, and (d) drinking coffee once a week or more often. An additional 35 participants were excluded for drinking coffee once or less often during the study period, and 4 additional participants were excluded for answering less than 50% of prompts. Thus, the final sample for analyses was slightly smaller than our preregistered target of 120.

To minimize attrition, compensation was linked to compliance. Paid participants received $20 for completing 80%-100% of EMA prompts, $15 for completing 50%-80%, and $5 for completing less than 50%. Participants who received course credit had a similar 3-tier compensation system.

***Procedure***

 **Intake Session**. After providing informed consent, participants reported their coffee drinking habit strength, coffee drinking intentions and attitudes, coffee drinking attributions, and demographics (see *measures* below). In addition, to obscure the purpose of the study and limit reactivity, participants answered an identical set of measures about soft drinks. Participants then wrote down implementation intentions (Adriaanse et al., 2011) to overcome potential obstacles for completing the prompts (e.g., “if my phone beeps when I am with people, then I will excuse myself and answer the prompt”).

 **Ecological Momentary Assessment (EMA).** For five weekdays (i.e., participants were not prompted on Saturday and Sunday), participants were prompted to respond 8 times per day at regular two-hour intervals from 8am to 10pm. Each prompt included items meant to obscure the purpose of the study, including location (e.g., home, campus) and temperature (hot, cold, or comfortable). Participants then reported how tired they were, whether they drank coffee in the past two hours, and whether they drank soft drinks in the past two hours. Participants also completed an exploratory mood item and an open-response item in which they briefly described their current situation.

 In addition, because fatigue on waking up may be particularly important for coffee drinking, participants completed a prompt every morning when they got out of bed. Morning prompts included the same items as the regular prompts as well as an item asking whether they had already drunk coffee (a measure of compliance). Thus, we could measure the prospective effect of waking fatigue on coffee drinking and avoid the self-report bias that might emerge with concurrent reports (i.e., “I drank coffee, therefore I must have been tired”).

 At the end of the first study day, participants with response rates of 50% or above (4 or more prompts) were informed of their approximate level of compliance via email (50-75% or 75-100%). Those with less than 50% compliance were contacted by phone and/or text message to address potential technical difficulties that might have led to low compliance.

 **Final Survey.** Participants were sent the final survey on the weekend after they completed the EMA portion of the study. This survey included the context-specific habit measure, single-event self-attribution measure, and open-text measures asking about self-regulation and general study feedback.

 **Follow-Up Survey.** Shortly after all data collection was completed, participants were emailed a survey that included the incentivized self-attribution measure (see below).

***Measures: Intake***

Additional measures for this study are included in the supplemental materials.

**Habit Strength.**

***Behavior-Frequency-In-Context (BFiC;*** Galla & Duckworth, 2015; Ji & Wood, 2007). Participants reported how often they drink coffee on a scale ranging from 1 (*Less than once a week*) to 5 (*More than 7 times a week; that is, more than once a day*). They then rated how often they drink coffee at the same time of day, and at the same location, on scales ranging from 1 (*Never or almost never at the same* [*time / location*]) to 5 (*Almost always or always at the same* [*time / location*]). Each participants’ coffee drinking frequency rating was then multiplied by the time and location stability ratings separately, and the two Frequency X Context scores were averaged to create a mean habit strength score.

***Self-Report Habit Index*** (SRHI; Verplanken & Orbell, 2003; This is the complete version of the brief 4-item SRBAI used in Study 1; Note that we did not include the full SRHI in Study 1 to minimize participant burden)**.** Participants indicated their agreement with a set of eleven statements regarding coffee drinking (e.g. “drinking coffee is something that I do without thinking,” “drinking coffee is something that belongs to my daily routine”), on a scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*).

 **Self-Attribution.**On scales ranging from 0 – 100%, participants rated the extent to which their coffee drinking was driven by habit (“my past behavior and habits”) and by fatigue (“my energy levels and tiredness”). Anchors of 0% indicated that coffee drinking was unaffected by a factor and 100% that coffee drinking was completely determined by a factor. Participants were instructed to not allow the sum of both ratings to exceed 100%.

 **Coffee Drinking Intentions and Attitudes.** On a scale ranging from 0 (*not at all*) to 100 (*extremely*), participants rated their liking of coffee drinking (“how much do you enjoy drinking coffee?”). On a scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*), participants rated their coffee drinking intentions (“I intend to drink coffee \_,” with the underscore replaced by the individual participant’s self-reported frequency of coffee drinking; Ajzen, 2002).

***Measures: Ecological Momentary Assessment (EMA)***

 **Fatigue.** Participants rated how tired they were on a scale ranging from 1 (*not at all*) to 6 (*extremely*).

**Coffee Drinking*.*** To ensure that a single coffee consumed over a period of time was categorized as one episode, participants indicated whether they started drinking coffee in the past two hours (*No* / *Yes - 1 Drink* / *Yes - 2 Drinks* / *Yes - 3 Drinks or more*). Answer choices were categorized into a binary drink/did not drink indicator of coffee drinking.

**Mood.** In this exploratory measure, participants rated their current mood on a scale ranging from 1 (*unhappy*) to 5 (*happy*).

 **Situation Description (Open-Text Measure).** In a free response, participants briefly described their current situation (e.g., “going to the gym,” “with friends”). Specifically, for prompts in which they indicated that they had recently been drinking coffee, they described that coffee drinking situation. For prompts in which they did not drink coffee, they described the situation they were in one hour previous. These situation descriptions were then used in the **“**Context-Specific Habit Measure”and the **“**Single-Event Attribution Measure” (see below).

***Measures: Final Survey***

 **Context-Specific Habit Measure.** As an exploratory measure, we randomly selected for each participant seven situation descriptions for prompts in which participants reported not drinking coffee and up to seven situation descriptions in which participants did report drinking coffee. For each situation description, participants rated (a) how often they drank coffee in that situation, (b) how automatic they perceived coffee drinking to be in that situation, and (c) the strength of their intentions to drink coffee in that situation (see Supplement for full description).

 **Single-Event Self-Attribution.** To evaluate attributions for a specific instance of a behavior, participants were shown their open-text situation description for their own final coffee drinking event, and rated the extent to which habit and fatigue contributed to drinking coffee at that time. Item wording and answer choices were the same as the intake attribution measure. To confirm that participants recalled the specific coffee drinking event, they reported whether they remembered it, and analyses for this single-event measure only included the 81 participants who answered affirmatively (72%).

***Measures: Follow-Up Survey***

**Incentivized Self-Attribution**. Participants were offered a monetary incentive of $3 if they accurately estimated the effects of fatigue and habit strength on their own coffee drinking during the study week. The incentive, along with using their own data as an objective benchmark, were designed to encourage participants to respond accurately and reduce any influences from social desirability or conversational norms. A total of 78 (70%) participants responded to the follow up survey and thus provided this rating.

**Results**

 Means, standard deviations, and between-person correlations for key variables appear in Table 3. The 112 participants (level 2 sample size) produced 3550 individual observations (level 1 sample size), corresponding to an average response rate of 31.7 out of 40 EMA prompts (79%).

On average, participants drank coffee a little over five times during the five-day period (*M* = 5.26, *SD* = 3.09), or approximately once a day. Scores on both habit strength measures suggested moderate coffee drinking habits. Furthermore, the two measures were strongly correlated with each other, *r* = .73, 95% *CI* [.63, .81]. Choice of habit measure did not have a noticeable impact on the results, and thus analyses are reported using the behavior-frequency-in-context scale (see table S2 in supplementary materials for analysis results using the SRHI).

***Primary Analyses***

Results were analyzed using the following multilevel model:

**Level 1:**

$$coffee\_{ij}=β\_{0j}+β\_{1j}fatigue\\_cmc\_{ij}+e\_{ij}$$

**Level 2:**

$$β\_{0j}=γ\_{00}+γ\_{01}habit\_{j}+γ\_{02}mean\\_fatigue\_{j}+γ\_{03}attribution\_{j}+u\_{0j}$$

$$β\_{1j}=γ\_{10}+γ\_{11}habit\_{j}+γ\_{12}attribution\_{j}+u\_{1j}$$

Where *i* and *j* represent observations (*i*) nested within persons (*j*); *coffee* refers to whether the participant did or did not report drinking coffee in the *following* prompt (i.e. a *lead* indicator of coffee drinking meant to capture the prospective association between fatigue and coffee drinking in the following two hours); This lagged design controls for response biases associated with concurrently reporting a predictor and outcome; *fatigue\_cmc* is a person mean-centered fatigue rating at each EMA prompt, computed by subtracting each participant’s mean level of fatigue from each fatigue rating, so that positive values reflect higher-than-average fatigue for that person and negative values reflect lower-than-average fatigue for that person; *mean\_fatigue* is each person’s mean level of fatigue, *habit* is each participant’s habit strength score, and *attribution* is each person’s attribution of coffee drinking to fatigue.

**Table 3**

*Study 2: Means (M), Standard Deviations (SD), and Between-Person Correlations*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***M (SD)*** | **1** | **2** | **3** | **4** | **5** | **6** |
|  |  |  |  |  |  |  |  |
| 1. Habit strength (BFiC) | 11.79 (5.21) |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |
| 2. Habit strength (SRHI) | 4.09 (1.35) | .73\*\* |   |   |   |   |   |
|   |   | [.63, .81] |   |   |   |   |   |
|   |   |   |   |   |   |   |   |
| 3. Mean fatigue | 3.44 (0.75) | -.05 | .09 |   |   |   |   |
|  |  | [-.23, .14] | [-.09, .28] |   |   |   |   |
|   |   |   |   |   |   |   |   |
| 4. Habit attribution | 36.99 (27.59) | .41\*\* | .46\*\* | -.07 |   |   |   |
|   |   | [.25, .56] | [.30, .60] | [-.25, .12] |   |   |   |
|   |   |   |   |   |   |   |   |
| 5. Fatigue attribution | 69.14 (23.43) | -.34\*\* | -.24\* | -.03 | -.37\*\* |   |   |
|   |   | [-.49, -.16] | [-.41, -.05] | [-.22, .16] | [-.52, -.20] |   |   |
|   |   |   |   |   |   |   |   |
| 6. Coffee counta | 5.26 (3.09) | .54\*\* | .44\*\* | -.08 | .20\* | -.14 |   |
|   |   | [.39, .66] | [.28, .58] | [-.26, .11] | [.01, .37] | [-.32, .04] |   |
|   |   |   |   |   |   |   |   |
| 7. Response rate | 31.70 (4.42) | -.05 | -.06 | -.01 | -.16 | -.13 | .03 |
|   |   | [-.23, .14] | [-.24, .13] | [-.19, .18] | [-.33, .03] | [-.31, .06] | [-.16, .21] |
|   |   |   |   |   |   |   |   |

*Note.* Scores for the BFiC ranged from 1-5 and SRHI ranged from 1-7, with higher numbers reflecting stronger habits. Mean fatigue ranged from 1-6, with higher numbers reflecting higher fatigue. Habit attribution and mood attribution ranged from 1-100, with higher scores reflecting stronger attributions. Response rates reflect the number of prompts answered (out of 40 possible), with higher scores indicating higher response rate. Values in square brackets represent 95% confidence intervals.

a Total number of coffee drinking events reported during the study period.

\* *p* < .05

\*\* *p* < .01

***Habit and Fatigue Effects on Coffee Drinking - Actual***

Model estimates for the primary multilevel model are shown in Table 4. To facilitate interpretation and reduce multicollinearity, all predictors in all regression analyses below were standardized to have a mean of 0 and standard deviation of 1. Due to convergence issues with the original frequentist model, we re-specified the main model as Bayesian. To avoid imposing restrictive priors on the results, we specified uninformative priors for all model predictors (a prior slope value of 0 with a standard deviation of 100).

To test whether habit determined coffee drinking as well as or better than fatigue, we compared the standardized coefficients for habit strength and person-mean centered fatigue ($γ\_{01}$and $γ\_{10}$in the model). As anticipated, participants with stronger habits were more likely to drink coffee, *OR*= 1.35, 95% CI [1.16, 1.55]. Yet, participants also drank more coffee when fatigued (within-person), *OR* = 1.22, 95% *CI* [1.08, 1.39]. *CI*s for all other model effects spanned 1.00 (see Table 4).

**Table 4**

*Study 2: Coefficient Estimates for Fixed Effects in the Main Multilevel Model*

|  |  |  |
| --- | --- | --- |
| **Variable** |  ***Odds Ratio*** |  **95% CI** |
|  |  |  |
| 1. Habit strength $γ01$ | 1.35 | 1.16-1.55 |
|   |  |  |
| 2. Fatigue (person-mean centered)$ γ00$ | 1.22 | 1.08-1.39 |
|   |  |  |
|   |  |  |
| 3. Fatigue (average person-level) $γ02$ | 0.96 | 0.84-1.09 |
|  |  |  |
|   |  |  |
| 4. Fatigue attribution $γ03$ | 1.02 | 0.89-1.18 |
|   |  |  |
|   |  |  |
| 5. Habit Strength X Fatigue (person-mean centered) interaction $γ11$ | 0.95 | 0.84-1.07 |
|   |  |  |
|   |  |  |
| 6. Fatigue (person-mean centered) X Fatigue Attribution interaction $γ12$ | 0.96 | 0.86-1.09 |
|   |  |  |

*Note*. The 95% CI *(Credible Interval*) represents the range of values that has a 95% chance of including the population odds ratio.

**Figure 1: Study 2**

*Likelihood of Drinking Coffee as a Function of Habit Strength and Fatigue*

****

*Note***.**  Percentages reflect the likelihood of coffee drinking by the following EMA prompt as a function of amount of fatigue in the prior prompt and strength of participants’ habit to drink coffee. Fatigue scores represent each participant’s mean +/- 1 standard deviation.

To determine whether these effects held also for the first coffee of the day, we computed a separate multilevel analysis that predicted coffee drinking on the first scheduled prompt of each day (i.e., excluding the participant-initiated morning prompts) from waking fatigue as measured in morning prompts, habit strength, and an interaction between habit and waking fatigue. The final sample size for this analysis consisted of 100 participants (level-2 sample size) and 307 responses (level-1 sample size, corresponding to participant-days). Out of the original 498 responses, 54 were excluded because participants reported that they already drank coffee by the time that they completed the prompt, and an additional 145 morning reports were excluded because they were submitted after the first regular EMA prompt of that day, and therefore could not be used to predict drinking in that prompt. Early morning fatigue was unrelated to coffee drinking by the following prompt, *OR*  = 0.96, 95% *CI* [0.65, 1.42], *p* = .843. Participants with stronger habits were more likely to drink first thing in the morning, *OR*  = 2.19, 95% *CI* [1.38, 3.47], *p* < .001. The interaction between habit strength and fatigue was not significant, *p* > .2. Thus, consistent with our expectations, waking fatigue did not influence coffee drinking on the first prompt of the day, whereas habit strength did.

Given that the present study measured rather than manipulated habit, we examined whether habit uniquely determined coffee drinking over and above the contribution of attitudes or intentions concerning coffee drinking. When these two predictors were added to the main model, the results remained essentially unchanged, with significant main effects of habit strength and fatigue but no effects for liking or intentions (see Table S3 in supplement). Thus, consistent with our hypotheses, the effect of habit strength on coffee drinking was not due to liking for coffee or intentions to drink it.

***Habit and Fatigue Effects on Coffee Drinking - Attributed***

To test our hypothesis that participants would attribute coffee drinking to fatigue more than habit, we computed a paired-samples *t*-test comparing the within-person difference between each participant’s fatigue and habit attributions. As expected, fatigue attributions were significantly stronger than habit attributions *M*difference = 32.14, 95% CI [24,18, 40.11], *t*(110) = 8.00, *p* < .001, *d* = 1.26, 95% CI [0.84, 1.67].

To ensure that the attribution results were not due to a failure to recall coffee drinking events or to ambiguity of attributions for multiple instances of a behavior, we evaluated the single-event attribution measure for participants’ last coffee drinking episode. Consistent with hypotheses, a paired-samples *t*-test revealed that participants attributed their most recent coffee drinking event to fatigue more than habit, *M*difference = 25.86, 95% CI [15.65, 36.08], *t*(80) = 5.04, *p* < .001, *d* = 0.94, 95% CI [0.50, 1.38].

Suggesting that the attribution findings are not due to social desirability or conversational norms, a paired samples *t*-test with the incentivized self-attribution measure, designed to maximize accuracy, revealed significantly stronger fatigue attributions than habit attributions, *M*difference = 16.01, 95% CI [5.82, 26.20], *t*(77) = 3.13, *p* = .002, *d* = 0.62, 95% CI [0.19, 1.04]. Thus, incentivized participants still overwhelmingly rated fatigue as more important than habit, even though incentives reduced the size of this effect (a difference of about 16 when incentivized compared with about 32 in the non-incentivized measure).

***Alternative Habit Measure: Effects of Context-Specific Habit -- Actual***

The final survey assessed a novel, within-person measure of habit strength to compare with the within-person measure of fatigue. A multilevel model predicted actual coffee drinking at each prompt (yes/no) from fatigue (within-person) and context-specific habit (within-person). As anticipated, context-specific frequency of coffee drinking was an especially strong determinant of coffee drinking, *OR* = 1.86, 95%CI[1.52, 2.32]. As in the main analysis, fatigue predicted coffee drinking as well, *OR* = 1.24, 95% CI[1.05, 1.49]. The larger effect of habit compared with fatigue, as well as the non-overlapping credible intervals, reveal that this measure of habit exerted a stronger effect on coffee drinking than fatigue.

***Correspondence Between Perceived and Actual Effects***

 This study’s repeated longitudinal design allowed us to estimate not only the overall effect of fatigue, but also whether the actual within-person effect of fatigue on coffee drinking was related to attribution to fatigue. That is, we assessed whether those who actually drank more in response to fatigue were aware of this relation in that they were more likely to make fatigue attributions. If participants’ attributions to fatigue are based on shared cultural theories rather than personal experience, however, we should find that attributions to fatigue are unrelated to the actual within-person effect of fatigue on coffee drinking. For this analysis, our main multilevel model tested whether attribution to fatigue moderated the within-person lagged association between fatigue and coffee drinking ($γ12$ in the model). A positive slope would suggest that participants with stronger fatigue-coffee drinking associations also gave stronger fatigue attributions (i.e., more accurate attributions). Supporting our hypothesis, strength of attribution to fatigue did not moderate the association between within-participant fatigue and coffee drinking at the next prompt, *OR* = 0.95, 95% CI [0.82, 1.11]. Thus, participants strongly attributed to fatigue regardless of its actual effect on their own coffee drinking, consistent with the notion that attributions draw on shared cultural lay theories (e.g., Wilson et al., 1982).

***Exploratory Analyses - Intensity of Experience and Attributions***

 Correlational analyses provided additional insight into the accuracy of participants’ attributions (see Table 3). First, attributions to fatigue were not correlated with mean fatigue levels, *r*(109) = -.03. This lack of effect is consistent with the weak correlations found in Study 1 between mood intensity and attributions, along with prior findings that attributions often reflect shared cultural lay theories more than individual experience (Wilson et al., 1982).

Second, on a correlational basis, participants with stronger coffee habits made stronger habit attributions, *r*(109) = .41 (behavior frequency in context), *r*(108) = .46 (self-report habit index). Thus, showing some evidence for accuracy, participants with stronger habits attributed their coffee drinking more to habit. To probe this effect, we divided our sample into tertiles by habit strength (weak, moderate, and strong). To enable us to compare the attribution measure with the actual influences on coffee drinking, we computed our main multilevel model predicting coffee drinking by extracting within-person effects for fatigue and habit for each participant (using the context-specific habit measure; to match the attribution measure, model slopes in log-odds units were standardized to have a range of 0-100%). Weak habit participants attributed on average a 51% difference in favor of fatigue, compared with an actual difference of 18% in the opposite direction (i.e., in favor of habit). Participants with moderate habits attributed a 39% difference in favor of fatigue, compared with a 21% difference in favor of habit. Strong habit participants attributed 4% in favor of fatigue, compared with an actual difference of 33% in favor of habit. Thus, participants with stronger habits correctly made stronger habit attributions and weaker fatigue attributions, but they continued to favor fatigue more than was merited by the actual predictors of their behavior.

***Exploratory Analyses - Downstream Effects of Attribution Accuracy on Well-Being***

To identify downstream effects of attributions, we assessed whether attribution to habit over fatigue is associated with a more positive mood in life, as measured using the average of participants’ mood reports in the ecological momentary assessment prompts. Attribution scores were calculated as the difference between each participant’s attributions to mood and to habit, so that positive scores implied attribution to fatigue more than habit, and negative scores implied attribution to habit more than fatigue. Attribution scores were moderately and negatively correlated with mood, *r*(109) = -0.27, 95% CI [-0.44, -0.09], so that attributing to habit over fatigue was associated with more positive mood.

To explore whether attributions favoring habit exert a unique effect on mood, or whether this correlation is simply due to people who are less tired and have stronger habits being more happy, we fit a linear regression model predicting mood from participants’ habit strength (measured using the behavior-frequency-in-context), mean fatigue levels (using the ecological momentary assessment), and attribution difference scores. Supporting a unique effect of attribution, stronger attribution to habit over fatigue predicted more positive mood, *b* = -0.12, 95% CI [-0.23, -0.02], *β* = -0.23. Lower mean fatigue levels were also associated with more positive mood, *b* = -0.22, 95% CI [-0.31, -0.13], *β* = -0.40. Habit strength did not show a discernible association with mood, *b* = 0.04, 95% CI [-0.06, 0.14], *β* = 0.04. Thus, these exploratory analyses suggest that greater recognition of habit in one’s own behavior is associated with higher well-being.

**Discussion**

In this second study, participants explained the causes of a mundane everyday action—coffee drinking—and then tracked their momentary fatigue and coffee drinking over the course of a typical week. Again, our hypotheses compared the actual influences on behavior to participants’ behavioral attributions. Fatigue and habit strength had comparable effects on actual behavior. The strong effect of habit maintained across three different measures of habit. Furthermore, analyses on the first coffee drink of the day and the within-person habit measure supported our hypothesis that participants would drink in response to habit more than fatigue. If participants’ attributions were accurate, they should have featured habit as much or more than fatigue. However, participants miscalibrated these behavioral influences by attributing their coffee drinking primarily to fatigue rather than habit.

Notably, Study 2 revealed a bias to overlook habit despite design features to reduce misattribution. The inaccuracy in self-attribution persisted when participants were incentivized for accuracy or asked at the end of the study about a recent coffee drinking event rather than their coffee drinking in general. Thus, it emerged despite motivation, an objective criterion, and, at least for our sample of coffee drinkers, adequate opportunity to observe their behavior given the frequency of consumption (once a day on average). Further attesting to the robustness of this attribution bias, coffee drinking is a mundane, everyday action that, unlike our first study, is not commonly preceded by a salient emotion-inducing experience.

**General Discussion**

 In two studies, participants’ attributions overemphasized inner states and underemphasized habit. Participants’ actual willingness to donate time in a laboratory task as well as their everyday coffee drinking were determined as much or more by habits than by inner states (mood and fatigue, respectively). However, participants’ attributions for why they acted the way they did emphasized inner states more than habit. Thus, participants appear to be both undervaluing habit compared with its actual influence on behavior and overvaluing inner states such as mood and fatigue. This pattern is understandable given the disproportionate value people place on personal introspections (Pronin, 2009) as well as general information- and motivation-based tendencies to interpret actions as goal-directed (Rosset, 2008). Through these forces, people may form socially-shared lay theories about behavior that inform their attributions. This *lure of phenomenology* not only biases lay theories but also may have oriented psychological theories to overvalue salient, motivational determinants of behavior (Duckworth et al., 2016).

 The combination of experimental manipulation in Study 1 and naturalistic observation in Study 2 provides evidence for the causal role of habits as well as the relevance of this attribution bias in everyday settings. Furthermore, the results replicated across the different measures of habit strength appropriate in these different tasks: Study 1’s manipulation of practice along with a reaction time measure; Study 2’s self-report measures of behavioral repetition in a given context (a determinant of habit formation) and experienced automaticity (a consequence of habit formation); and Study 2’s exploratory within-person, context-specific habit measure tapping participants’ history of repetition in specific situations.

A number of features of our research would be expected to maximize participants’ accuracy and minimize biases. Each study assessed attributions for specific recent behaviors, minimizing biased recall. Furthermore, in both studies, attributional biases were evident even when participants were incentivized for accuracy, as well as when participants were explaining others’ behavior (Study 1) and regardless of whether participants were explaining a repeated behavior in general or a specific recent instance (Study 2). These patterns did not stem from difficulties in interpreting the habit items or plausibly applying them to behavior, as our participants made habit attributions for their own and other’s behavior even when accuracy was incentivized. Additionally, in pretesting for Study 2, habit was identified as one of the top reasons for drinking coffee.

That our participants discounted habit influences on their own behavior may seem at odds with the actor-observer effect, which proposes that people have a bias to attribute their own behavior more to environmental factors that they do others’ behavior (Jones & Nisbett, 1971). Technically, habits are not solely an environmental influence, as they reside both in a person’s learned associations in memory and in the environment that triggers a habitual response (Wood & Rünger, 2016). And yet, even if one considers habits an environmental force, a meta-analysis of the literature showed that the actor-observer effect on attributions emerges largely under specific conditions, such as for negative events rather than positive ones (Malle, 2006). The behaviors examined in this paper were presumably either positive (helping) or neutral (drinking coffee). It may be that people would more readily implicate habits in driving undesirable behaviors—the proverbial “bad habits.”Relatedly, both studies recruited fluent English speakers, predominantly residing in the U.S. and U.K. If habit underestimation depends on agency beliefs, it may be smaller in collectivistic cultures that place less emphasis on individuals and more emphasis on context (e.g. Crandall et al., 2001).

 Similar to other lay theories, habit underestimation may have important downstream effects. For example, it raises questions about the accuracy of people’s reports that lack of willpower is the primary reason for their failures to lose weight, save money, and exercise (American Psychological Association, 2012). It also raises questions about the effectiveness of common self-regulation strategies: If people misattribute the sources of their behavior, then they may focus on strategies that affect inner states (e.g. reduce coffee drinking by reducing fatigue) at the expense of situational strategies that may more successfully modify habits. This would align with the argument that situational self-regulation strategies are relatively non-salient, which could lead people to overlook these interventions’ potential (Duckworth et al., 2016). Suggesting that accurate attributions are beneficial, exploratory analyses in Study 2 revealed that participants who placed more weight on habit in their attributions also reported more positive mood. It may be that well-being increases not only with habit performance (Heintzelman & King, 2019) but also with recognizing habits’ elusive yet pervasive role in daily life.

**References**

Adriaanse, M. A., Vinkers, C. D. W., De Ridder, D. T. D., Hox, J. J., & De Wit, J. B. F. (2011). Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. *Appetite*, *56*, 183–193. <https://doi.org/10.1016/j.appet.2010.10.012>

Ajzen, I. (2002). *Constructing a TpB Questionnaire: Conceptual and Methodological Considerations*. Undefined.<https://www.semanticscholar.org/paper/Constructing-a-TpB-Questionnaire%3A-Conceptual-and-Ajzen/6074b33b529ea56c175095872fa40798f8141867?p2df>

American Psychological Association. (2012). *Stress in America: Our Health at Risk*.

Avni‐Babad, D. (2011). Routine and feelings of safety, confidence, and well-being. *British Journal of Psychology*, *102*(2), 223–244. <https://doi.org/10.1348/000712610X513617>

Crandall, C. S., D’Anello, S., Sakalli, N., Lazarus, E., Nejtardt, G. W., & Feather, N. T. (2001). An attribution-value model of prejudice: Anti-fat attitudes in six nations. *Personality and Social Psychology Bulletin*, *27*(1), 30–37.<https://doi.org/10.1177/0146167201271003>

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science*, *11*, 35–55. <https://doi.org/10.1177/1745691615623247>

Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of Personality and Social Psychology*, *109*(3), 508–525.<http://dx.doi.org.libproxy1.usc.edu/10.1037/pspp0000026>

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G. J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 102. <https://doi.org/10.1186/1479-5868-9-102>

Gershman, S. J., Gerstenberg, T., Baker, C. L., & Cushman, F. A. (2016). Plans, habits, and theory of mind. *PLoS ONE*, *11*(9).<https://doi.org/10.1371/journal.pone.0162246>

Hagger, M. S., Rebar, A. L., Mullan, B., Lipp, O. V., & Chatzisarantis, N. L. D. (2015). The subjective experience of habit captured by self-report indexes may lead to inaccuracies in the measurement of habitual action. *Health Psychology Review*, *9*(3), 296–302.<https://doi.org/10.1080/17437199.2014.959728>

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour, 3*(12), 1252-1262. <https://doi.org/10.1038/s41562-019-0725-0>

Heintzelman, S. J., & King, L. A. (2019). Routines and meaning in life. *Personality and Social Psychology Bulletin, 45*(5), 688-699. <https://doi.org/10.1177/0146167218795133>

Ji, M. F., & Wood, W. (2007). Purchase and consumption habits: Not necessarily what you intend. *Journal of Consumer Psychology*, *17*, 261–276. <https://doi.org/10.1080/10577400701542403>

Knowlton, B. J., & Diedrichsen, J. (2018). Editorial overview: Habits and skills. *Current Opinion in Behavioral Sciences, 20,* iv-vi. <https://doi.org/10.1016/j.cobeha.2018.02.009>

Mazar, A., & Wood, W. (2018). Defining habit in psychology. In B. Verplanken (Ed.), *The psychology of habit: Theory, mechanisms, change, and contexts* (pp. 13–29). <https://doi.org/10.1007/978-3-319-97529-0_2>

[Mills, C., & D’Mello, S. (2014). On the validity of the autobiographical emotional memory task for emotion induction. *PLOS ONE*, *9*(4), e95837. https://doi.org/10.1371/journal.pone.0095837](https://www.zotero.org/google-docs/?RrpM7d)

McFerran, B., & Mukhopadhyay, A. (2013). Lay theories of obesity predict actual body mass. *Psychological Science, 24*(8), 1428–1436. https://doi.org/10.1177/0956797612473121

Pronin, E. (2009). The introspection illusion. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 1–67). [https://doi.org/10.1016/S0065-2601(08)00401-2](https://doi.org/10.1016/S0065-2601%2808%2900401-2)

Rosset, E. (2008). It’s no accident: Our bias for intentional explanations. *Cognition, 108*(3), 771-780. <https://doi.org/10.1016/j.cognition.2008.07.001>

[Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242.](https://www.zotero.org/google-docs/?RrpM7d) <https://doi.org/10.1177/0022022106297301>

[Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, *33*(6), 1313–1330.](https://www.zotero.org/google-docs/?RrpM7d) <https://doi.org/10.1111/j.1559-1816.2003.tb01951.x>

Wen, C. K. F., , Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research*, *19*(4), 1–1.<https://doi.org/10.2196/jmir.6641>

Wilson, T. D., Laser, P. S., & Stone, J. I. (1982). Judging the predictors of one’s own mood: Accuracy and the use of shared theories. *Journal of Experimental Social Psychology*, *18*(6), 537–556. [https://doi.org/10.1016/0022-1031(82)90072-5](https://doi.org/10.1016/0022-1031%2882%2990072-5)

Wood, W., Mazar, A., & Neal, D. (2021). Habits and goals in human behavior: Separate but interacting systems. *Perspectives on Psychological Science.*

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*, 1281–1297. <http://dx.doi.org.libproxy2.usc.edu/10.1037/0022-3514.83.6.1281>

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314.<https://doi.org/10.1146/annurev-psych-122414-033417>

[Zhang, Z., & Yuan, K.-H. (2018). *Practical Statistical Power Analysis Using Webpower and R*.](https://www.zotero.org/google-docs/?RrpM7d)

1. We also computed an exploratory model that included the Habit x Mood interaction, which essentially replicated the results reported in the text: The hypothesized effect of habit, *OR* = 0.71, 95% CI[0.54, 0.92], *p* = .011, no effect of the sad condition, *OR* = 0.84, 95% CI[0.59, 1.21], *p* = .348, or happy condition, *OR* = 0.96, 95% CI[0.66, 1.39], *p* = .826, and only an unexpected nonsignificant interaction between habit and the happy mood condition, *OR* = 1.38, 95% CI[0.96, 1.99], *p* = .078. [↑](#footnote-ref-1)