To prevent means to know:

Explicit but no implicit agency for prevention behavior

Supplementary Material

Main experiment: Data selection and preprocessing

As per our pre-registration, participants were removed from the analyses if their data did not contain any baseline trials with keypresses or came with less than 10 trials in any design cell for the analysis of temporal binding. Trials with errors were excluded from all further analyses, comprising early and late keypresses (3.11% and 1.41%, respectively), and trials in which participants entered an estimate exceeding 60 minutes (2.03%), as well as trials after agency ratings. For the remaining trials, we computed estimation errors as the estimated time entered by the participants minus the actual time shown on the virtual clock at the time of the keypress. These estimation errors were then converted to milliseconds for display purposes.

In addition we performed an outlier correction by excluding estimation errors that deviated more than 2.5 standard deviations from a participant's cell mean (2.03% of the trials). The decision to include this correction was based on previous observations of a minority of extreme outliers at or close to the theoretical maximum error of 30 minutes (180°) on the virtual clock. We preferred to report the data after removing such outliers even though the observed pattern of results does not change qualitatively when including all trials in the analyses.

Temporal binding (action binding) was determined for each participant by subtracting the mean estimation error of the baseline condition from the mean estimation error of the operant and the prevention condition, respectively.

Main experiment: Additional methodological considerations

Typical studies on temporal binding (e.g., Haggard et al., 2002) only employ a rotating clock hand as visual stimulus but no additional stimuli such as the loading bar for timing potential actions. Even though this feature is not necessary to observe the reported pattern of results as documented by the supplementary experiments (see below), we chose to add this feature to rule out one potential concern. Without a precise timing of when each event in a trial would occur or be omitted, the operant and prevention condition would differ substantially regarding the variability of action-effect delays. That is: In operant condition a keypress action was followed by a tone with a constant delay of 300 ms. Because the interval between an action and its consequence can affect temporal binding (Ruess et al., 2017; Wen et al., 2015), the prevention condition requires an identical delay between the action and the omission of the to-be-prevented event to allow for a fair comparison. A similar method was used in previous work to study temporal binding for intended non-actions (Weller et al., 2020, Exp. 3). Crucially, this study found intact effect binding (i.e., a subjective pre-dating of the effect) if an event occurred because participants refrained from performing an action. This pattern suggests that temporal binding can be assessed robustly with the chosen paradigm while at the same time addressing potential alternative explanations in terms of action-effect delays if a condition does not yield evidence for temporal binding.

Main experiment: Follow-up analyses

In a set of post-hoc analyses we assessed whether block order modulated the pattern observed for temporal binding to gauge whether experiencing either the operant or the prevention condition in the first block would introduce crosstalk during later stages of the experiment (O-B-P-O-B-P or P-B-O-P-B-O, with O = operant, B =

baseline, P = prevention). Neither the perceptual shifts of the operant condition differed between both block orders as assessed with a *t*-test for independent samples, $t(91) = 1.38$, $p = .171$ (Levene test for homogeneity of variances: $p = .406$), nor did the perceptual shifts in the prevention condition, $t(91) = 0.41$, $p = .681$ (Levene test: $p = .687$).

Experiment S1: Aversive tone without bar

Apparatus and procedure

Experiment 1 was pre-registered on the Open Science Framework (https://osf.io/4usb8) and all procedures were as in the main experiment with the following exceptions. Crucially, the display only featured a clock face (diameter: 6.5 cm) with a rotating clock hand (frequency: one rotation per 2000 ms) but there was no loading bar. Participants went through 3 different conditions (operant, prevention, baseline), presented in separate blocks of 40 trials in a full within-subjects design.

Each trial started with the rotation of the clock hand. The starting point of the clock hand was marked by a white dot which helped to estimate the duration of a whole turn. In every condition, participants had to wait during the first turn of the clock hand, during the second turn they could freely choose whether to press the response key (action) or not (omission) – a keypress within the first turn or after the end of the second turn led to error messages. An action or omission was followed by different effects depending on the condition. In the *baseline condition* neither actions nor omissions triggered any additional effects. In the *operant condition* each keypress caused a pure sinusoidal tone (600 Hz) with a constant action-effect delay of 300 ms, whereas there was no tone if the participant chose not to press the response key. In the *prevention condition* white noise was played automatically at the end of the clock hand's second turn. Participants could prevent this stimulus by pressing the response key within the second turn. When the participant had pressed a key, the clock hand

further rotated for a random duration between 2000 and 3000 ms in every condition and stopped afterward.

Participants

Forty-eight participants were recruited for the study (mean age: 24.23 years; 34 females). All participants gave written informed consent and they received payment or course credit for their participation. Two participants had to be excluded because of *a priori* defined criteria (see the preregistration).

Results

Figure S1 shows the results of Experiment S1. As for the main experiment, we excluded trials with errors from all further analyses, comprising early and late keypresses (0.97% and 1.21%, respectively), and trials in which participants entered an estimate exceeding 60 minutes (0.13%), as well as trials after agency ratings. Of the remaining trials, 2.09% were removed as outliers (including these trials again did not affect the reported pattern of results).

Temporal binding and explicit ratings

The operant condition showed a reliable action binding effect (12.41ms), $t(45) = 3.85$, $p < .001$, $d_z = 0.57$, whereas the prevention condition did not, $t(45) = 1.33$, $p = .192$, $d_z = 0.20$. Follow-up analyses of Bayes Factors supported these findings by indicating strong support for the presence of an effect in the operant condition, $BF_{10} = 62.08$, and evidence for the absence of an effect in the prevention condition, $BF_{10} = 0.27$ ($BF_{01} = 3.72$; again computed as JZS Bayes Factors with a non-directional test using a Cauchy prior with a scale parameter of 1). A direct comparison of both conditions revealed a significant difference in action binding ($\Delta = 8.54$ ms), $t(45) = 2.45$, $p = .018$, $d_z = 0.36$, suggesting a marked difference in implicit measures of agency between operant and prevention behavior.

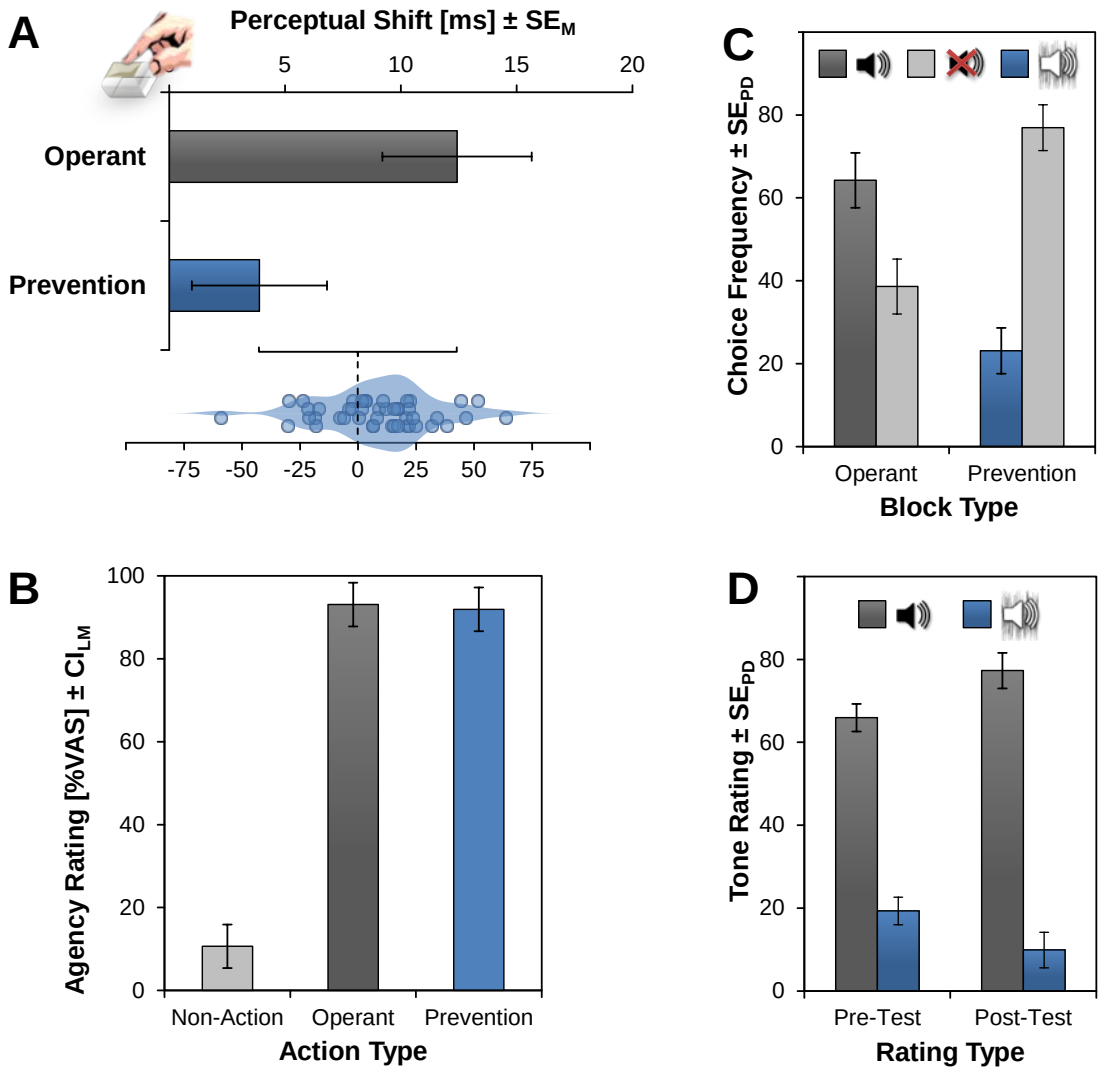

Figure S1. Results of Experiment S1. **(A)** Temporal binding for operant actions and prevention behavior. Binding scores were computed by comparing interval estimates in the prevention and the operant condition to the baseline condition (SE_M = standard error of the mean). Bayesian analyses indicated substantial evidence for the absence of an effect in the case of prevention behavior (BF_{01} = 3.72). The scatter plot shows difference scores between both conditions for each participant and an estimate of the corresponding density function. **(B)** Mean agency ratings for baseline, operant and prevention trials (CI_{LM} = 95% confidence intervals according to the method of Loftus & Masson, 1994). VAS = visual analogue scale. **(C)** Choice frequencies (in %) for operant and prevention blocks. Operant actions triggered a neutral tone (dark grey bar) whereas action omissions in prevention blocks were followed by white noise (blue bar). No sound events occurred in the remaining conditions (light grey bars). Error bars indicate standard errors of paired differences (SE_{PD}; Pfister & Janczyk, 2013) calculated separately for each block type. **(D)** Tone ratings on a VAS (0 = negative, 100 = positive) before and after the experiment (pre-test vs. post-test). Dark grey bars show mean ratings for the neutral tone whereas blue bars show mean ratings for white noise. Error bars were calculated separately for pre- and post-test.

Explicit agency ratings, by contrast, were equally high in the operant condition and the prevention condition, $t(45) = 0.54$, $p = .590$, $d_z = 0.08$, $BF_{10} = 0.13$ ($BF_{01} = 7.51$). Both conditions differed markedly from the baseline ps $< .001$, $d_z > 4.32$, $BFs_{10} > 8.20 \times 10^{23}$ as also reflected in an omnibus analysis of variance (ANOVA) across all three conditions, $F(2, 90) = 549.27$, $p < .001$, $\eta_p^2 = .92$ (corrected according to Greenhouse and Geisser's method; $\varepsilon = .798$), suggesting high subjective agency in operant and prevention behavior alike.

Manipulation Check

In the operant condition participants chose to act in 64%, suggesting a choice bias towards acting (and/or towards producing a neutral tone effect), $t(45) = 4.31$, $p < .001$, $d_z = 0.64$. As in the main experiment, the bias towards acting was even stronger in the prevention condition (77%), $t(45) = 9.75$, $p < .001$, $d_z = 1.44$, and a comparison of both conditions again suggested a significant difference, $t(45) = 3.94$, $p < .001$, $d_z = 0.58$. This pattern was again mirrored in subjective ratings of the tone valence which participants gave at the beginning (pre-test) and at the end of the experiment (post-test). A 2 (tone: neutral vs. white noise) x 2 (position: pre-test vs. post-test) repeated-measures ANOVA indicated that the white noise was rated as negative at both time points, $F(1, 45) = 294.66$, $p < .001$, $\eta_p^2 = .87$, while this effect increased with time, $F(1, 45) = 29.00$, $p < .001$, $\eta_p^2 = .39$ (main effect of position: $F(1, 45) = 0.20$, $p = .655$, $\eta_p^2 = .004$).

Experiment S2: Neutral tones

Apparatus and procedure

All settings were as for Experiment S1 with the following modifications. Experiment S2 only used two block types, a baseline block and an experimental block which contained operant and prevention trials alike. To this end, participants now

operated two response keys rather than the single response key of the previous experiment.

Each trial started with the rotation of the clock hand. In every condition, participants had to wait during the first turn of the clock hand, during the second turn they could freely choose whether to press any one of the two keys (action) or not (omission) – a keypress within the first turn (2.03%) or after the end of the second turn (2.08%) led to error messages. An action or omission was followed by different effects depending on the condition. In *baseline* blocks neither actions nor omissions triggered any additional effects. In experimental blocks, omissions always resulted in an automatic tone that would appear at the end of the clock's second turn. Pressing one of the keys would produce a different tone with a constant action-effect delay of 300 ms (*operant action*) whereas pressing the other key would prevent any tone from being played (*prevention action*). Both tones were neutral sinusoidal tones of distinct pitch (300 and 600 Hz) and a duration of 100 ms. Participants were explicitly informed about the condition of the upcoming block and the corresponding consequences that their choices would entail. The mapping of response keys to action type (operant vs. prevention) was constant for each participant but counterbalanced across participants. Each condition was presented in three blocks of 40 trials, with block order E-B-E-B-E-B or B-E-B-E-B-E counterbalanced across participants (E = experimental, B = baseline). Explicit agency ratings were only collected in the experimental block but for all of the three options (omission, operant, prevention). Participants were not asked to rate tone valence as neither tone came with aversive qualities.

Participants

Thirty-two participants (mean age: 26.38 years; 24 females) were recruited for the study. All participants gave written informed consent and they received payment or course credit for their participation. Four participants had to be excluded because of *a priori* defined criteria as for the main experiment.

Results

Figure S2 shows temporal binding, agency ratings and choice frequencies for Experiment S2. As in the preceding experiments, trials with errors were excluded from all further analyses, comprising early and late keypresses (2.03% and 2.08%, respectively), and trials in which participants entered an estimate exceeding 60 minutes (0.16%), as well as trials after agency ratings and outliers (2.21%; including outlier trials again did not affect the observed pattern of results).

The operant trials showed a reliable action binding effect (15.42ms), $t(27) = 2.94$, $p = .007$, $d_z = 0.56$, whereas the prevention trials did not, $t(27) = 0.26$, $p = .801$, $d_z = 0.05$. Follow-up analyses of Bayes Factors supported these findings by indicating support for the presence of an effect in the operant trials, $BF_{10} = 5.62$, and evidence for the absence of an effect in the prevention condition, $BF_{10} = 0.14$ ($BF_{01} = 6.98$). A direct comparison of both conditions revealed a significant difference in action binding (Δ = 14.20ms), $t(27) = 2.74$, $p = .011$, $d_z = 0.52$, again attesting a marked difference in implicit measures of agency between operant and prevention behavior.

Explicit agency ratings, showed different levels of agency for operant actions and prevention actions, $t(27) = 2.35$, $p = .026$, $d_z = 0.44$, $BF_{10} = 0.60$ ($BF_{01} = 1.67$). Both conditions also differed from the omission condition, in which participants did not prevent an automatic sound effect ps $< .001$, $d_z > 0.76$, $BF_{10} > 74.87$, as also reflected in an omnibus ANOVA across all three conditions, $F(2, 54) = 19.78$, $p < .001$, $\eta_p^2 = .42$ (corrected according to Greenhouse and Geisser's method; $\varepsilon = .698$).

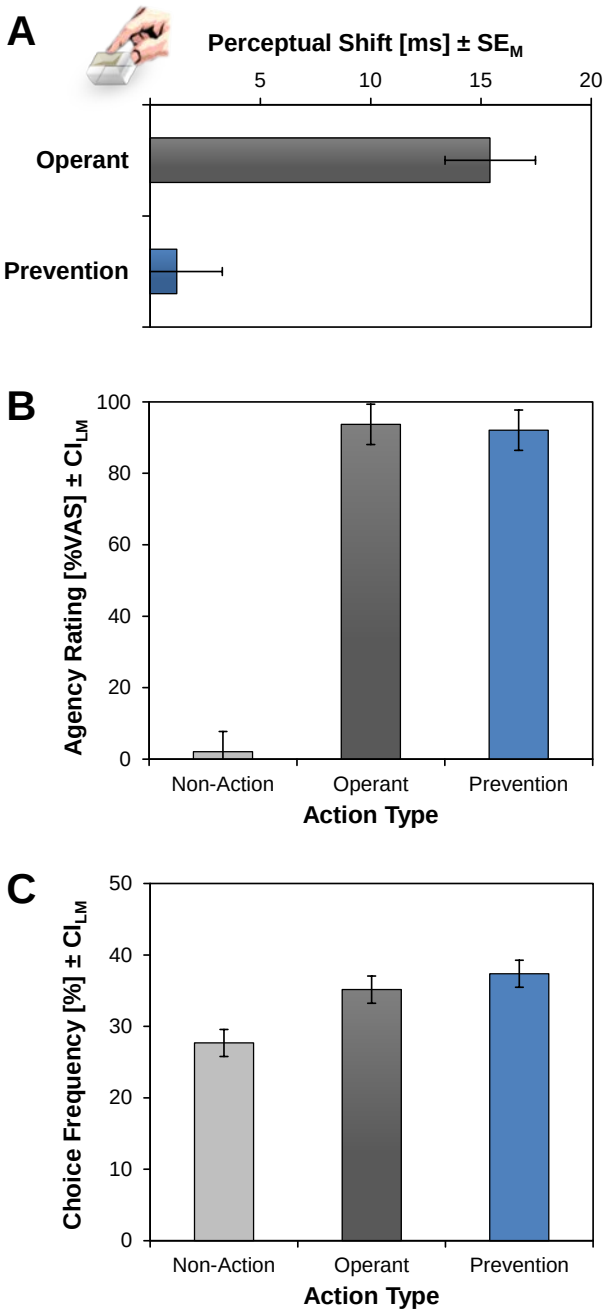

Figure S2. Results of Experiment S2. **(A)** Temporal binding for operant actions and prevention behavior. Binding scores were computed by comparing interval estimates for prevention and operant actions to the baseline condition (SE_M = standard error of the mean). Bayesian analyses indicated substantial evidence for the absence of an effect in the case of prevention behavior (BF_{01} = 6.64). **(B)** Mean agency ratings for non-action, operant and prevention trials of the experimental condition (CI_{LM} = 95% confidence intervals according to the method of Loftus & Masson, 1994). VAS = visual analogue scale. **(C)** Choice frequencies for non-action, operant and prevention trials of the experimental condition.