SUPPLEMENTARY MATERIAL

Feeling free: external influences on endogenous behavior

Lucie CHARLES a\*, Patrick HAGGARD a

#  SUPplementary METHODS

## Stimuli

Stimuli consisted of random-dot kinematograms (RDK) appearing in a 5°-diameter aperture presented on a computer monitor with a frame rate of 60 Hz. Each set of dots was shown for one video frame and then replotted three video frames later (Roitman and Shadlen, 2002, https://shadlenlab.columbia.edu/resources/VCRDM.html). The subset of dot moving coherently were replotted with an offset in the up or down direction from their original location, creating apparent upward or downward motion while the remaining dots were relocated randomly. This created an apparent motion of 5 deg/s. Coherence of the motion (i.e., proportion of dots moving in the same direction) were determined during the pre-test phase (see Procedure).

## Procedure

The experiment consisted of three distinct parts: a pre-test, the main experiment, and a post-test and was conducted in two sessions of 80-90 minutes. During session one, participants performed the pre-test (25 minutes), then 7 blocks of 96 trials of the main experiment followed by the post-test (5 minutes). During session 2, they performed 9 blocks of the main experiment followed by the post-test.

The goal of the pre-test phase was to establish for each participant the two coherence thresholds so that participants were either close to chance (Hard condition) or significantly above chance (Easy condition) to recognize bidirectional from the unidirectional motion. To determine those thresholds, participants were presented on each trials with three RDK stimuli positioned at the left (7.9 deg), center and right (7.9 deg) of the screen. Among the three stimuli, one of the stimulus was bidirectional while the two others were both unidirectional, going either both in the upward or downward direction, or one downward and the other upward. Participants performed an odd-one-out task, reporting which of the three stimuli they thought was bidirectional. Coherence of the stimuli was determined by two staircase procedures, aiming at reaching chance-level accuracy (33%) for the Hard condition and above-chance accuracy (50%) for the Easy condition. For the hard condition, the staircase followed a 3Down–1Up rule starting from coherence 5% and with a constant step-size of 1%. The rule was that coherence was decreased one step if accuracy over the last three trials was above 66%, remained constant for accuracy at 33% and increased one step if accuracy was at 0%. Therefore, if accuracy for a given coherence level was close to chance level, the coherence was likely to stay constant or slightly increase, while above-chance accuracy would lead to a decrease in the coherence value. For the easy condition, a staircase starting at 60% coherence with constant step-size of 1% was used. Computing accuracy over the past four trials, coherence was decreased of one step if accuracy was above 75%, remained constant if accuracy was at 50% and was increased of one step if accuracy was below 25%. The pre-test was stopped when targeted accuracy was reached over the last 40 trials for both staircases.

The goal of the post-test was to determine whether at the end of the experiment, participants were still at chance to discriminate the bidirectional motion from the unidirectional motion in the Hard condition and whether they were still above chance in Easy conditions. It consisted of an odd-one-out task similar to the pre-test consisting of 72 trials, intermixing the motion coherence levels of the Easy condition and the Hard condition. A chi-test was used to estimate whether each participants was at chance in the Hard condition and above-chance in the Easy condition. Data from both sessions were concatenated and participants not reaching the required accuracies were further excluded from analysis.

## Statistical analysis

Mean accuracy was computed both for the pre- and post-test for each difficulty condition. Note that chance level for pre- and post-test were distinct from the main experiment, corresponding to 1/3 = 33%. Mean accuracy was computed for the main experiment for Instructed trials and each difficulty condition where chance level was 50%. Accuracy values in the pre- and post-test and in the main experiment were submitted to two-tailed t-test to determine whether they differed from chance and from each other. Mean RT and mean Freedom of Choice ratings were computed for each type of trials (Free and Instructed) and for each difficulty condition (Easy and Hard trials) and entered into a repeated measure ANOVA with both elements as within-subject factors. T-tests were performed to determine differences between conditions (two-tailed, unless specified otherwise).

Regression were performed to test correlation between RT and Freedom of choice ratings. We used trial-based linear mixed-effects model with random intercepts for each participant to estimate the significance of the correlation between the two variables.

## Motion Energy Analysis

To do so, we use spatial filter applied on the trial frame-to-frame images of the dot-motion stimuli. We used a similar method than the one described in Kiani et al. (2008), using two pairs of spatiotemporal filters selective respectively for upward or downward movement (Adelson and Bergen, 1985; Kiani et al., 2008). Each filter is built based on an impulse function selective for either space or time, identical to Kiani (2008). The spatial filters correspond to even and odd symmetric fourth-order Cauchy functions (with parameters σg = 0.05° and σc = 0.5°), while the temporal impulse correspond to a linear filter (with time constant k = 60). By combining these filters together and summing pairs of filters, it is possible to create filters selective for a specific spatiotemporal frequency consistent with those of MT neurons. The filters were convolved with the three-dimensional spatiotemporal pattern of the RDK. The results of the two convolutions were then squared and summed together to obtain the local motion energies at each point in time. To calculate the net motion energy over the entire stimulus, local motion energies were summated across space too. Finally, as our stimulus contained only motion in two possible directions, motion energies for opponent directions were subtracted from each other to obtain the net motion energy in the stimulus. We choose arbitrarily to assign the upward direction a positive value, while downward motion was assigned a negative value.

As trials were associated with different motion coherence values in different conditions and for different participants, we expressed the motion coherence at each moment of each trial as a standardized difference from that participant’s mean motion coherence across all trials (normalization step, Figure 2B). We computed the mean and standard deviation over all the trials separately for each direction of motion, each condition (easy and hard) and each participant and used them to normalize each individual time-course accordingly. As the spatiotemporal filters, similarly to a Fourier transform, create distortions at the edge of the time-courses, we had to remove trials with very short or very long RT that would increase the noise in the overall signal from the computation. Therefore, we kept only trials with RTs between 500ms and 4500ms to perform the rest of the analysis.

However, for information we display below the signal before normalization both for instructed and free trials for each participant, in order to provide an estimate of the raw signal presented. For each trial, the sign of the time-courses was flipped according to the response made by the participants rather than a specific direction of the motion in the stimulus, so that positive values now indicated the choice made by the participant on that trial while a negative sign corresponded to the opposite choice. In order to obtain a measure of how much the noise fluctuation overall favored the decision made by the participant, we computed for each trial the cumulative sum or area under-curve (AUC) of each time-series over the whole decision time. The average values of the obtained trial-by-trial AUC values were average separately for each condition and each participant.



Figure S1: Mean and standard deviation of the AUC signals before normalization for each participant (x-axis) for the Easy (blue) and Hard (red) condition, for Instructed (left column) and Free (right column) trial. Circles represent mean signal across trials and lines represent standard deviations across trials.

In order to determine the whether participants were able to introspect whether their choice was guided by sensory noise, we tested whether a correlation was observed between the subjective reports of freedom of choice and the objective measure of how much participants were actually influenced by sensory information. To do so, we tested whether our measure of objective freedom from stimulus-determination (trial-by-trial AUC values) predicted the subjective freedom of choice ratings, using a linear mixed-model approach with random intercepts and slope for each participant, including RT as an additional predictor. We also conducted individual regressions for each participants and reported the obtained beta value in Figure 5.

# SUPplementary RESults

## Accuracy in pre-test, main experiment and post-test

The experiment started with a pre-test to determine the values of motion-coherence of the random-dot stimuli for the easy and hard conditions. The goal of this step was to find for each participant a set motion coherence values so that he or she would perform close to chance (hard condition) or above chance (easy condition) in identifying the bidirectional stimuli that cued free choices. This was required so that we could be sure, in the main experiment, that participants were either more or less aware that they were required to make free-choice decisions. Participants performed an identification task, finding the one stimulus that had bidirectional motion of random dots in the context of two other stimuli consisting of either upward or downward motions. Two interleaved staircase-procedures were used to find a coherence level at which participants were close to chance level (hard condition, staircased accuracy at 33%, red lines in Figure S2A), and another coherence level at which they were clearly above chance (easy condition, staircased accuracy arbitrarily at 50%, blue lines in Figure 2A) in identifying which of the three stimuli was bidirectional. As can be seen in Figure2A, our manipulation was successful in finding those thresholds.

In order to verify that these difficulty levels yielded the expected level of accuracy thorough the experiment, we performed a post-test session at the end of each main experiment session. Participants performed the same odd-one out task as during the pre-test session (see methods), using the estimated easy and hard coherence thresholds (Figure S2C). Accuracy for one participant in the hard condition exceeded chance level, and he was therefore excluded from further analysis. Five further participants had accuracy that did not differ statistically from chance in the easy condition, when tested in both post-test sessions separately. For those participants, average accuracy in the main task in easy trials was close to 70%, compared to ~90% for the rest of the participants. It suggests that for those participants, the staircase procedure was unsuccessful to determine the appropriate motion coherence level, most likely because participants’ accuracy decreased during the actual experiment compared to the pre-test phase. They were therefore also excluded from further analysis. For the remaining 14 participants, we found that in the post-test session, they were still no better than chance in detecting bidirectional free cues in the hard condition (t(13) = -1.17, p = .26, d = -0.31), while they remained above chance in the Easy condition (t(13) = 7.38, p < .001, d = 1.97). This confirmed that the two conditions elicited different awareness of the decision-related stimuli.

During the main experiment, subjects were presented with a central random-dot kinematogram whose motion was either upward or downward (Instructed trials) or a mixture of both directions (Free trials). In Instructed trials, subjects were told to press the up or down key if they thought the motion was directed upward or downward, respectively. In Free trials, when the motion was bidirectional, they were free to press either of the two keys. Hard and easy trials were randomly intermixed, using the coherence values determined in the pre-test phase, so that in half the trials participants could not identify whether they were in an instructed or free trial, while in the remaining half decision context was clearly identifiable.

Accuracy in easy instructed trials was unsurprisingly above chance level (Figure S2B t(13) = 12.6, p < .001, d = 3.37) and exceeded that of hard instructed trials (t(13) = 11.9, p < .001, d = 3.17). Participants also performed slightly above chance when they judged the direction of unidirectional motion on instructed hard trials (t(13) = 4.91, p < .001, d = 1.31). It suggests that even in the hard condition, participants were able to extract some information about the upward or downward direction of the motion. Taken together, these results suggest that our staircase procedure was successful in finding thresholds such that participants could clearly identify trials as free or instructed in the easy but not in the hard condition. Importantly, accuracy was between chance and ceiling in both conditions.

****

*Figure S2: Accuracy in the Pre-test, the main Experiment and the post-test. A: Plot of the cumulative accuracy across trials for the Easy (blue lines) and Hard (red lines) for each participant during the odd-one-out task of the pre-test (chance level 33%). B: Accuracy in instructed trials during the main experiment for the easy (blue bar) and hard condition (red bar) averaged across participants (chance level = 50%). C: Accuracy in the odd-one-out task of the post-test post session for the easy (blue bar) and hard condition (red bar) averaged across participants (chance level = 33%). Error bars indicate standard error of the mean.*

## Effect of trial type and difficulty on RT, Freedom of choice and response bias

Next, we investigated how trial difficulty and trial type (Free vs Instructed) influenced the subjective ratings of freedom of choice, response-times and response bias.

Consistent with the high accuracy observed in the easy instructed condition, these trials were also associated with faster response-times (RT) than other conditions (Instructed Hard: t(13) = -4.56, p < .001, d = -1.22; Free Hard: t(13) = -4.52, p < .001, d = -1.21; Free Easy: t(13) = -5.38, p < .001, d = -1.44) confirming that they corresponded to an easy perceptual choice. No other significant differences in RT were observed in the remaining conditions (Free Easy vs Hard: t(13) = -1.48, p = .16, d = -0.40; Hard Instructed vs Free: t(13) = -1.02, p = .33, d = -0.27; Free Easy vs Instructed Hard: t(13) = -1.16, p = .27, d = -0.31), which resulted in a significant interaction between difficulty and trial-type (F(1,13) = 23.4, p < .001, η² = 0.64).

Despite no difference in RTs being observed between easy and hard free trials, we cannot exclude different factors contributed to increasing or decreasing RTs in each of the condition. While in the easy trials participants might take more time to confirm they identified bidirectional motion, in hard trials they might on the opposite “give up” more easily to detect the direction motion and respond faster.

****

*Figure S3: Response-times (RT) and subjective ratings of freedom of choice. A: RT averaged across participants for each difficulty condition and trial-type (Instructed versus Free). B-C: Histograms of RT averaged across participants for each difficulty (line) and trial-type (column) condition.*

We next considered whether RT and freedom of choice were associated on a trial-by-trial basis. Separate linear mixed models were fitted for each level of perceptual difficulty and each condition. We found a positive relation between the two variables in both easy and hard trials in instructed and in free-choice conditions (Instructed Easy: F(1,5278) = 377, p < .001; Instructed Hard: F(1,5278) = 330, p < .001; Free Easy: F(1,5278) = 310, p < .001; Free Hard: F(1,5278) = 230, p < .001), suggesting that trials with longer decision-times were associated with stronger feelings of free choice. To determine whether this correlation varied across condition, we fitted a linear model separately for each participant and the regression coefficients were analysed using a repeated-measures ANOVA with difficulty and trial-type as factors. No significant main effect or interaction between factors were found, suggesting a global relation between RT on freedom of choice ratings, independent of our experimental design. We nonetheless included RT as a covariate in further analyses of subjective freedom of choice, to remove the variance in subjective experience attributable to performance factors alone.

## Bias of previous response on action selection and Freedom of choice

Finally, we explored in more details the pattern of responses in free trials. In particular, our goal was to determine if participants had consistent biases towards one or the other response hand and how such biases related to RT. Overall, percentage of left-hand responses did not differ from 50% across participants neither in the easy condition nor in the hard condition. Participants’ bias was overall consistent between the easy and hard condition and distributed evenly between left and right hand biases (Table 1). Importantly, no clear pattern between strength of bias and RT was detected.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 |
| Easy | % Left Responses | 0.33 | 0.27 | 0.32 | 0.40 | 0.41 | 0.40 | 0.34 | 0.50 | 0.51 | 0.58 | 0.60 | 0.50 | 0.61 | 0.66 |
| RT (ms) | 792 | 2699 | 2638 | 1814 | 2409 | 2280 | 2611 | 1585 | 1533 | 1854 | 1397 | 1803 | 2274 | 2512 |
| Hard | % Left Responses | 0.34 | 0.42 | 0.38 | 0.30 | 0.42 | 0.44 | 0.59 | 0.47 | 0.47 | 0.52 | 0.56 | 0.69 | 0.60 | 0.75 |
| RT (ms) | 855 | 3099 | 2728 | 1976 | 2067 | 2464 | 2924 | 1594 | 1889 | 1771 | 1456 | 1332 | 2166 | 3123 |

*Table S1: Response bias and RT in free trials for the Easy and Hard condition across participants (column). Bias is presented as the % of left hand Responses. The participants’ data are sorted according to the strength of the bias, from minimum values corresponding to an overall right hand bias, to maximum values representing to a left hand bias.*

To further investigate how participants selected a response in Free trials, we investigated whether previous response biased the following response on the next trial. To do so, we computed the proportion of repeat and switch responses in Free trials preceded by Instructed (Figure S4A) or Free (Figure S4B) trials as well as the associated changes in Freedom of choice ratings (Figure S4C-D).

In hard trials, we observed a systematic bias toward repeating the same response whether the preceding trial was Instructed (Figure S4A) or Free (Figure S4B) and Easy or Hard (Previous trial Instructed Easy: t(13) = 4.63, p < .001, d = 1.24; Previous trial Instructed Hard: t(13) = 5.64, p < .001, d = 1.51; Previous trial Free Easy t(13) = 4.99, p < .001, d = 1.33; Previous trial Free Hard: t(13) = 7.02, p < .001, d = 1.88). In Easy trials, the bias towards repeating the same response was observed only when the preceding trial was a Free trial, irrespective of whether it was a Hard or Easy trial (Figure S4B; Previous trial Free Easy t(13) = 3.00 , p = .010, d = 0.80; Previous trial Free Hard t(13) = 2.17, p = .049, d = 0.58). No bias was observed in Easy trials preceded by Instructed trials (Figure S4A; Previous trial Instructed Easy t(13) = 1.23, p = .24, d = 0.33; Previous trial Instructed Hard: t(13) = 1.37, p = .19, d = 0.37), suggesting participants relied on a different signal to make a free choice in this context. Taken together, these results suggest that participants relied more heavily on previous motor signals when no evidence was present to guide the current choice, either because the stimulus was particularly unclear or either because they had to make several successive free choices.

Interestingly, despite observing a biasing effect of the previous response on the current trial, no difference in subjective freedom of choice was observed when repeating or switching response in any of the conditions (all p> 0.14), suggesting participants fail to introspect the biasing effect of the previous response.



*Figure S4: Proportion of response repetition (A-B) and average Freedom of choice after repeat or switch response (C-D) following an Instructed trial (A,C) or a Free trial (B,D) for Easy (blue) and Hard (red) conditions.*