**Supplemental Material**

**Valence ratings**

An Expression (angry, happy, surprised) x Filter (LSF, HSF) repeated measures ANOVA revealed a significant effect of Expression (F(2,104) = 2044.27, *p* < .001, η2 = .98), such that ratings were more negative for angry than surprised faces, which were more negative than happy faces (*p*s < .001). There was also a significant interaction (F(2,104) = 7.0, *p*= .001, η2 = .12), demonstrating the effect of Expression was evident for both LSF and HSF images (*p*s < .001). Further, HSF images were rated as more negative than LSFs only for happy faces (*p* = .002; angry and surprised *p*s > .3).

**Initiation and response times**

Consistent with the mouse-tracking analyses in the main text, we used generalized estimating equations (GEE) multi-level regressions to examine initiation times and response execution times.

*Initiation times.* We regressed initiation time (i.e., when participants initiate movement) onto Rating (positive, negative), Filter (LSF, HSF), and the interaction, which revealed a main effect of Filter (B=27.55, SE=2.87; Z=9.61; p<.001), with slower initiation times for HSF than LSF images. This is sensible, given that HSF images have long been known to induce more deliberate processing. The main effect of Rating was not significant (B=4.60, SE=3.38; Z=1.36; p=.17), nor was the interaction (B=3.07, SE=5.88; Z=0.52; p=.60).

*Response times.* Mouse-tracking studies have commonly found *x*-coordinate deviation to co-vary with response times (i.e., when participants select a response), as trajectories that deviate to a greater degree tend to take longer time to ultimately select a response. Using a multi-level GEE regression, across all trials, *x*-coordinate deviation was a strong positive predictor of response times (B=283.43, SE=15.05; Z=18.84; p<.001), consistent with considerable previous work (e.g., Freeman & Ambady, 2010; Stolier & Freeman, 2017).

Next, we regressed response times onto Rating (positive, negative), Filter (LSF, HSF), and the interaction, which revealed a main effect of Rating (B=44.53, SE=14.94; Z=2.98; p=.003), with response times slower for positive than negative trials, consistent with prior work (Neta et al., 2009; Neta & Tong, 2016). There was also a significant main effect of Filter (B=35.66, SE=7.57; Z=4.71; p<.001), such that response times were slower for HSF than LSF images, also as expected.

**Bootstrapping analyses**

Our analyses used GEE multi-level regression models at each of the 101 time steps for the four conditions: Rating (positive, negative) x Filter (HSF, LSF). At each time step, we regressed trajectories’ *x*-coordinates onto Rating (positive, negative), Filter, (HSF, LSF), and their interaction. Our primary hypothesis concerned the Rating x Filter interaction. To solve the issue of multiple statistical testing for this interaction at 101 time steps, we used a bootstrapping approach taken from Dale et al. (2007) except here adapted for our factorial design and our GEE regression framework. The bootstrapping results showed that the experiment-wide significance of the Rating x Filter interaction was guaranteed at a criterion of *p* < .05, *p* < .01, or *p* < .001 if a minimum of 5, 6, or 8 consecutive time-steps in the trajectory showed a significant interaction effect.

We performed a bootstrap of 10,000 simulated experiments, wherein 106 model participants (the *n* of the study) were constructed. Each participant’s trajectories for the four conditions were constructed using a normal distribution based on the 101 time-step means and standard deviations for that participant’s actual trajectory. That is, for each of the four conditions within each of the 106 model participants within each of the 10,000 simulations, each time step (out of 101) was sampled from a normal distribution using the mean and standard deviation associated with that time step in the participant’s actual mean trajectory. We then ran GEE multi-level regression models at each of the 101 time steps for each of the 10,000 simulations. Among these 10,000 simulations, we recorded the frequency with which consecutive sequences of time steps all showing a significant interaction effect (*p <* .05) occurred.

The simulations revealed that consecutive sequences of 5, 6, 7, and 8 time-steps (showing a significant interaction effect at *p* < .05) occurred with a frequency of 1.74%, 0.38%, 0.15%, and 0.03%, respectively. Thus, a sequence of 5 significant consecutive time-steps was produced by chance with less than a .05 probability (5%), a sequence of 6 or 7 significant consecutive time-steps produced by chance with less than a .01 probability (1%), and a sequence of 8 significant consecutive time-steps produced by chance with less than a .001 probability (0.1%). Thus, the GEE interaction effect had to be significant (*p* < .05) at a minimum of 8 consecutive time-steps across the trajectories to guarantee an experiment-wide alpha of .001.