Sensory and action neural tuning explains how priors guide human visual decisions

Gabriela Iwama[1,2,3] and Randolph F. Helfrich[1]

[1]Hertie Institute for Clinical Brain Research, Center for Neurology, University Medical Center Tübingen, Tübingen, Germany.
[2]University of Tübingen, Graduate Training Center of Neuroscience, Tübingen, Germany
[3]International Max Planck Research School for the Mechanisms of Mental Function and Dysfunction, Max Planck for Biological Cybernetics, Tübingen, Germany.

Correspondence

Gabriela Iwama and Randolph F. Helfrich, Hertie Institute for Clinical Brain Research, Center for Neurology, University Medical Center Tübingen, Otfried-Müller-Str. 27, 72076 Tübingen, Germany, phone: +49 7071 29 89189, fax: +49 7071 29 4488, E-Mail: gabriela.iwama@uni-tuebingen.de and randolph.helfrich@gmail.com

Conflict of interest

The authors declare no competing interests.

Author contributions

Conceptualization: GI, RFH; Methodology: GI, RFH; Formal Analysis: GI; Investigation: GI; Resources: RFH; Writing – Original Draft, GI; Writing – Review & Editing: RFH; Supervision: RFH; Funding Acquisition: GI, RFH.

Abstract

Prior expectations bias how we perceive the world. Despite well-characterized behavioral effects of priors, such as confirmation bias, their neural mechanisms remain unclear. Contemporary theories postulate conflicting predictions: does the brain enhance expected sensory information (sharpening the expected representation) or rather prioritize unexpected information (dampening the expected representations)? Here, we combined reversal learning with a noisy motion discrimination task to investigate how priors impact sensory and action information processing. Using behavioral modeling to infer participants' latent priors and EEG to track neural dynamics, we demonstrated that priors differentially impact sensory and action representations over time. While priors introduced long-lasting biases on action coding irrespective of their validity, sensory information was selectively enhanced with confirmatory evidence. Critically, dampening of action representations predicted confirmation biases, whereas sensory tuning dynamics tracked speed-accuracy trade-offs. These findings reveal a dissociable and temporally dynamic influence of priors in visual decisions, reconciling competing theories of predictive perception.

Keywords

Expectation, predictive processing, confirmation bias, speed-accuracy trade-off, belief updating, electroencephalography (EEG)

Significance statement

Prior expectations, abstract rules, and recent decisions impact perception and ensuing actions. While priors are beneficial in most circumstances, they might be detrimental in other scenarios, as exemplified by confirmation biases, serial response biases, or switch costs. Albeit well-characterized at the behavioral level, the neural mechanisms underlying these biases remain undetermined. By tracking participants' latent priors and neural dynamics with high-density EEG, the authors demonstrate that priors differentially impact sensory and action information, uncovering the neural mechanisms underlying confirmation biases and speed-accuracy trade-offs in predictive perception.

Introduction

In a dynamic and uncertain world, sensory evidence is often limited. Thus, the ability to flexibly make perceptual decisions based on available prior and current information is fundamental for survival. An inherent speed-accuracy trade-off arises for almost every perceptual decision: should one rely on prior knowledge or gather more evidence before committing to a choice? For example, prior information on which trees are likely to bear fruit can save time and energy if correct, but relying on it may incur delays and opportunity costs if the prior proves wrong, leading to confirmation and choice history biases[1–3]. While the effect of priors on behavior is well-characterized, it remains unclear how the brain flexibly integrates prior information with ongoing sensory input to guide adaptive behavior.

There are two main theories of how prior information is integrated with sensory input: Bayesian and cancellation theories. According to *Bayesian theories*[4–6], sensory information is biased relative to its prior probability. Therefore, they predict a relative enhancement of expected sensory input (sensory sharpening). In contrast, *cancellation theories* predict a prioritization of informative signals given the limited capacity of the brain[7,8]. Since unexpected sensory inputs are more informative than expected ones, cancellation theories predict a relative dampening of expected sensory inputs. Therefore, these theories have conflicting predictions of how sensory processing is modulated by prior information.

What are the possible solutions to these contradictions? One possibility is that action and sensory processes are modulated by distinct regimes. Bayesian explanations are highly prevalent within the perceptual domain[5,9,10], whereas cancellation theories are mostly prevalent within the action literature[7,11]. Another explanation is that the modulation of sensory information is a dynamic process that depends on the validity of the prior as evidence is accumulated over time. As proposed by the Opponent Process Theory[8], sensory information is initially sharpened by prior expectations and only dampened if a prediction error is signaled by a bottom-up sensory process. While both hypotheses have received empirical support, both have limited power to explain the balance between speed and accuracy in visual decisions. To resolve this debate, it is fundamental to

3

91   investigate both action and sensory processing when priors are confirmed or disconfirmed
92   by sensory information during rapid decision-making.

93   Here, we investigated how prior expectations modulate action and sensory
94   information during a cued motion discrimination task. Critically, the cue validity changed
95   across trials, and participants had to adaptively use the cue information, the prior, across
96   trials. Behavioral modeling was employed to estimate participants' prior strength, while
97   multivariate decoding was used to assess high-density electroencephalography (EEG) to
98   capture sensory and action information. Our results show that prior expectations
99   distinctively modulate action and sensory processing while driving a dynamic sensory
100 tuning that supports flexible decision-making.

101

Results

103 To assess how prior expectations modulate sensory information and ensuing actions,
104 participants completed a cued motion discrimination task (**Fig. 1a**). The cues indicated
105 which direction of motion was more likely to appear in the upcoming stimulus. Cue validity
106 was unknown to the participants. Each trial began with participants selecting one of two
107 cues (orange or blue) and rating their confidence in its validity (low or high). Next, the cue
108 was presented indicating a direction (left or right), followed by a Random Dot
109 Kinematogram (RDK) with titrated left/right motion coherence to achieve 75% accuracy.
110 Participants reported the perceived motion direction (left/right) on a response box. As
111 stimuli were titrated to perceptual threshold and included high noise levels, participants
112 were instructed to rely on the cue information (i.e., the prior) to guide their motion choices.

113 In this task design, we modulated the strength of prior information by varying cue
114 validity across trials. A block design alternated between informative and non-informative
115 trials. In informative blocks, one cue had a validity of 80%, while the other cue had a
116 validity of 30%. In non-informative blocks, both cues had low validity (30%). Prior to the
117 experiment, participants were informed that the validity of the cues would change but
118 were not informed of the timing or frequency of these changes. Therefore, participants
119 had to learn the true validity of the cues through experience.

4

120   To encourage participants to balance making accurate motion choices against

121   learning the cue validities, a dual reward system was implemented (see *Methods*).

122   Participants were rewarded for correct motion choices and for selecting a valid cue. Cue

123   rewards were scaled by their reported confidence. If participants reported high confidence

124   in their selected cue, they entered a high-risk, high-reward state. In contrast, reporting

125   low confidence placed participants in a safe reward scheme with lower potential gains

126   and substantially reduced potential losses. This design incentivized participants to report

127   low confidence when they were uncertain about the cue validity. However, this reward

128   system does not guarantee that participants will use the prior information during the

129   decision.

130   To promote a flexible reliance on prior information, we induced a speed-accuracy

131   trade-off. Participants were penalized if they did not respond to the perceived direction of

132   motion within 1000 ms of RDK onset. In this design, it is beneficial to rely on the prior

133   even when no coherent motion is perceived. To optimize rewards, participants had to

134   learn which cue was the best and adaptively use the prior information over trials.

135   A key feature of this task is that it separates the cue choice from the cue information and

136   the motion choice. This design disentangles the process of learning the prior (the cue

137   validity) from forming a specific visual prior (left or right motion) and making a perceptual

138   choice (left or right button response). As a result, it allows assessing how prior

139   expectations modulate both sensory (motion perception) and action (overt response)

140   information processing at different stages.

141

Latent priors track reported confidence and learning

142

143   Participants' latent priors (i.e., their subjective belief about the cue's validity at each trial)

144   were estimated using the Volatile Kalman Filter[12] (VKF; **Fig. 1b**), which models

145   environmental volatility (see *Methods*). The VKF was fitted to participants' cue choices

146   and outcomes using Laplace Approximation (see **Fig. 1c** for a single-subject example).

147   To validate the estimated priors, we compared how they varied depending on participants'

148   reported confidence, block type, and cue accuracy on informative blocks. Across

149   participants, prior strength was significantly higher when confidence was high (**Fig. 1d**; p

150   < .001, t(19) = 8.22, d = 1.84) and when the block was informative (p < .001, t(19) = 12.30,

151   d = 2.75). Moreover, participants more frequently selected the true best cue in an

152   informative block when priors were stronger (median split; p < .001, t(17) = 10.78, d =

153   2.54). Together, these findings provide robust evidence supporting the validity of the

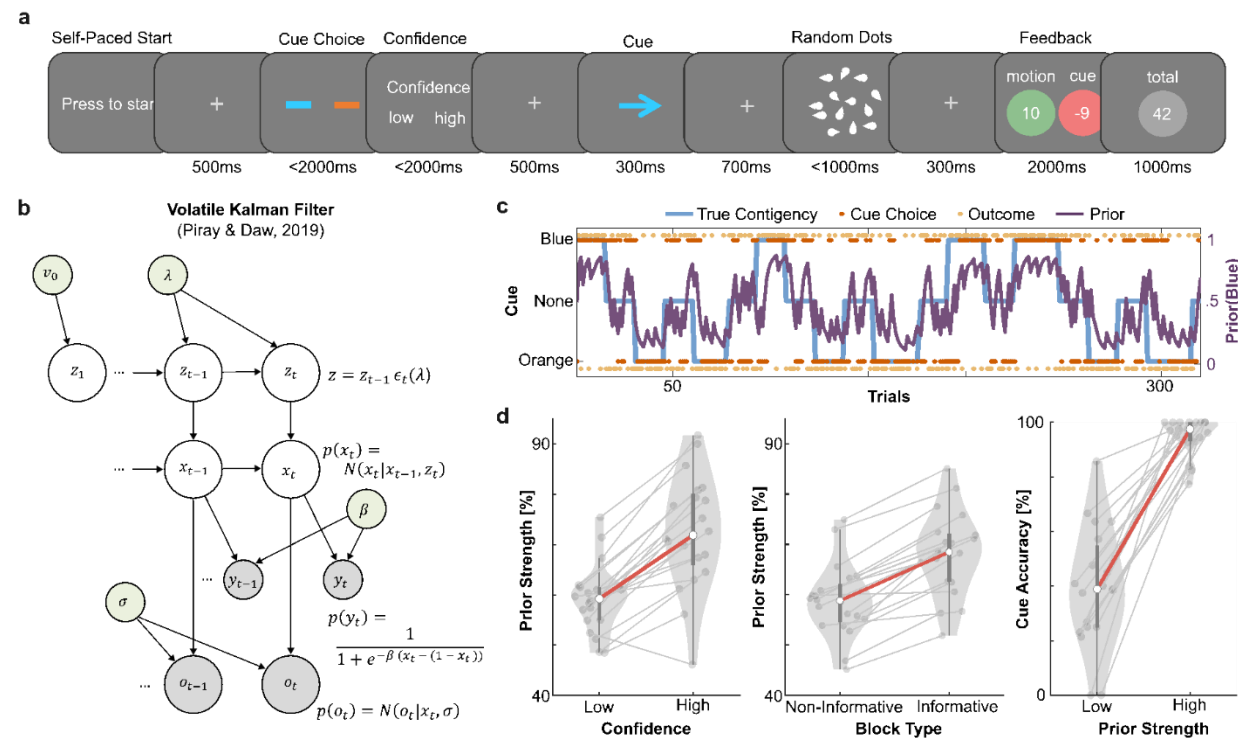154   priors—as estimated by the model—as a measure of participants' latent priors.

155

156   **Fig. 1 | Experimental design, modeling, and validation. a,** On every trial, participants selected either the
157   orange or the blue cue, rated their confidence (low or high) that the selected cue would be valid, and were
158   subsequently presented a cue (colored arrow) indicating the direction of the Random Dot Kinematogram
159   (RDK; left or right). After a delay, participants were presented with the RDK at their individual discrimination
160   level (threshold at 75% accuracy) and made a decision about the motion direction (left or right). At the end
161   of each trial, participants received feedback about their motion choice and the validity of the chosen cue.
162   **b,** The Volatile Kalman Filter generative model[12] contains two hidden Markov chains that dynamically
163   diffuse over time, in which x_t denotes whether the blue cue is the most informative, and z_t denotes the
164   inverse volatility of x_t. The outcome $o_t \in \{0,1\}$ consists of the observed cue validity, where 0 = blue invalid
165   or orange valid and 1 = blue valid or orange invalid. The choice, y_t, is given by a softmax function of x_t with
166   a free parameter, β, to model how rationally participants follow their prior estimates. The free parameters
167   λ, σ, and v_0 indicate the (inverse) volatility update rate, the observation noise, and the initial volatility. **c,**
168   Single subject example of best cue (true contingency; blue), cue choice (y_t, orange), outcome (o_t; yellow),
169   and estimated prior (x_t; purple) across trials. Note that the estimated prior closely tracked reversals in the
170   true contingency. **d,** Estimated latent prior of chosen cue according to confidence judgments (left), block
171   type (middle), and cue accuracy according to prior strength during informative blocks (right). Violin plots
172   show the median, interquartile range (IQR), and $\pm 1.5\ IQR$.

173

Priors increase the influence of cue information during motion decisions

Having validated the estimated prior, we examined whether participants integrated the prior into their visual decisions. We performed a trial-wise regression analysis on motion accuracy and reaction time, with prior strength and cue validity interaction as a predictor. Additionally, a random intercept for each participant was included to account for individual differences in performance. The interaction effect was statistically significant for both accuracy, (**Fig. 2a**; t(6004) = 10.19; f^2 = .019) and reaction time (**Fig. 2b**; p < .001; t(6004) = -7.19; f^2 = .011). Hence, the behavioral benefit of valid cues increased when the prior was high for both accuracy ($Benefit_{ACC} = ACC_{High,Valid} - ACC_{Low,Valid}$; $M_{benefit}$ = 4.1%, SD = 6.6%) and reaction time ($Benefit_{RT} = RT_{Low,Valid} - RT_{High,Valid}$; $M_{benefit}$ = 16 ms, SD = 27 ms). Conversely, invalid cues incurred an increased behavioral cost when the prior was high for both accuracy ($Cost_{ACC} = ACC_{High,Invalid} - ACC_{Low,Invalid}$; M_{cost} = 4.1%; SD = 9.1%) but not for reaction time ($Cost_{RT} = RT_{High,invalid} - RT_{Low,Invalid}$; M_{cost} = -7 ms; SD = 29 ms). These analyses indicate that participants adaptively used the cue information during their decision-making, relying on it more when their latent prior was high and sensory evidence was confirmatory. After validating that the task promoted the flexible use of priors in visual decisions across trials, we next examined individual differences in prior integration.

Individual differences reveal rigid and flexible prior integration

To test individual differences in how priors are integrated, we correlated the costs, benefits, and validity modulation in both accuracy and reaction time. First, when priors were strong, the modulation of validity in accuracy ($\Delta ACC = ACC_{High,Valid} - ACC_{High,Invalid}$) was negatively associated with the modulation of reaction time (**Fig. 2c**; $\Delta RT = RT_{High,Valid} - RT_{High,Invalid}$; r = -.79, p < .001). This indicates that participants who were more biased to judge the motion direction toward the cue also responded faster when the cue was valid. We will denote the variance along this shared dimension as *confirmation bias*. Additionally, the costs in accuracy ($Cost_{ACC}$) were positively correlated with the reaction time benefits (**Fig. 2d**; $Benefit_{RT}$; r = .67, p < .001), reflecting individual differences in the speed-accuracy trade-off. Specifically, participants who made faster

7

204 decisions when the cue was valid made more mistakes when the cue was invalid. The
205 variance along this shared dimension will be denoted as *speed-accuracy trade-off*. In
206 sum, these correlations indicate that participants employ distinct strategies to integrate
207 priors into visual decisions. This raises the question of *whether* and *how* these strategies
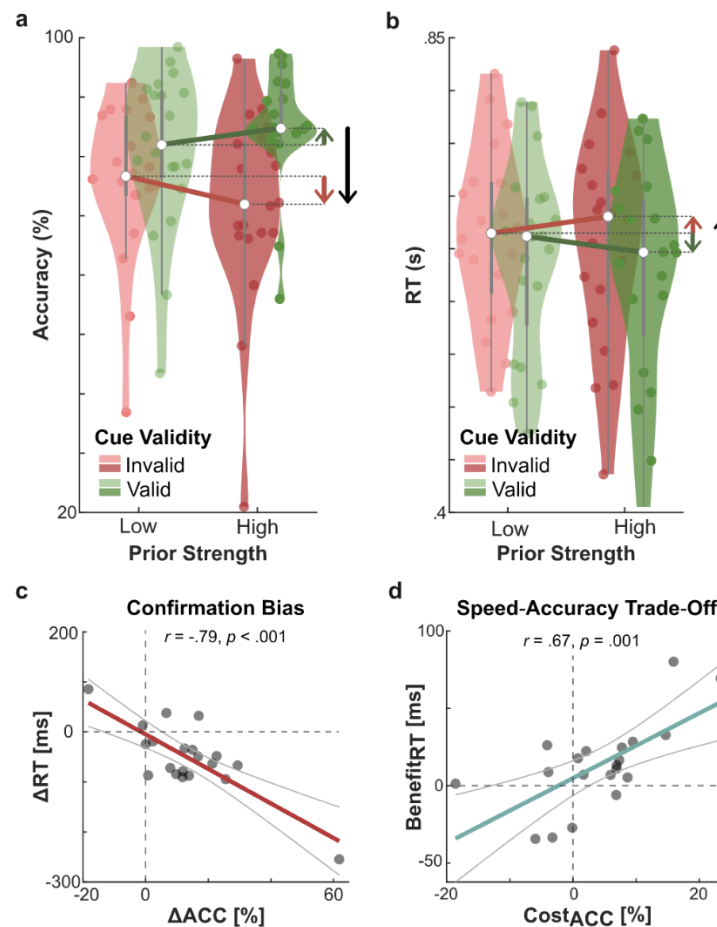208 are associated with sensory and action representations.

209

210

211 **Fig. 2 | Priors induce confirmation biases and speed-accuracy trade-offs. a,** Accuracy (ACC) of motion
212 choice as a function of cue validity and prior strength (subject-level median split into high and low trials).
213 Green and red arrows illustrate the group-level behavioral benefit ($Benefit_{ACC}$) and cost ($Cost_{ACC}$) of high
214 prior when compared to low prior trials within the same cue validity, respectively. Black arrows indicate the
215 behavioral modulation of high prior trials depending on cue validity (ΔACC). **b,** Same conventions as in
216 panel (**a**) but for reaction time (RT). **c,** Correlation between the difference in accuracy for high prior trials
217 when the cue was valid and invalid ($\Delta ACC = ACC_{High,Valid} - ACC_{High,Invalid}$) and the difference in reaction
218 time for the same conditions (ΔRT, excluding incorrect trials). **d,** Correlation between the benefit in reaction
219 time ($Benefit_{RT} = RT_{Low,Valid} - RT_{High,Valid}$) and the cost in accuracy ($Cost_{ACC} = ACC_{Low,Invalid} -$
220 $ACC_{High,Invalid}$).

221

Priors distinctively stabilize sensory and action information

To investigate how priors affect sensory and action information during visual decision-making, both uni- and multivariate analyses of high-density EEG recordings were employed. To investigate whether priors modulate the total activation of channels, a cluster-based permutation test across channels and time was used to compare the event-related potential (ERP) of the data time-locked to the RDK onset and reaction time (RT) for each prior strength (high, low) and cue validity (valid, invalid) condition (see *Supplementary Figure 1*). However, no statistically significant clusters across time and channels were observed before or after RDK onset when comparing high versus low prior in each validity condition. Therefore, a multivariate decoding approach was applied to test if priors affect channel covariance.

Before assessing how priors modulate sensory and action information, we first decoded both the sensory (left or right motion) and the action information (left or right button response) from high-density EEG in a time-resolved manner (**Fig. 3a/b**). All decoding analyses relied on a Linear Discriminant Analysis[13], a dimensionality reduction method that finds a linear combination of features (channels) to maximize the separation between classes (left/right motion or choice) while minimizing the within-class variance. To avoid training biases, we employed a 10-fold cross-validation with balanced training samples regarding the decoded feature (action or sensory), the prior strength (high, low), and the cue validity (valid, invalid).

Across participants, single-trial true class probabilities (TCP) of testing samples were significantly decoded above chance after stimulus presentation for sensory (p = .041, $\bar{t}_{cluster}$ = 2.77; d = .78; cluster-based permutation test) and action information (p = .001, $\bar{t}_{cluster}$ = 3.48; d = 1.25; cluster-based permutation test), as well as reaction time ($p_{sensory}$ = .004, $\bar{t}_{cluster}$ = 3.40; d = .95; p_{action} = .001, $\bar{t}_{cluster}$ = 4.59; d = 1.62; cluster-based permutation test). Additionally, for action, we observed a cluster even before the RDK onset, [-.020 s .086 s], p = .047, $\bar{t}_{cluster}$ = 2.44, d = .63; cluster-based permutation test), suggesting that the cue already influences the action information before sensory processing.

Having demonstrated that sensory and action information can be decoded, we tested whether prior strength modulates sensory or action information. TCP differences

9

253   between weak and strong priors were compared for each validity condition. As no
254   statistically significant differences were found (see *Supplementary Figure 2*), we
255   determined whether prior strength modulates the temporal evolution of sensory and
256   action information over the trial. To examine how sensory and action information evolve
257   over time, we assessed the temporal cross-generalization of the decoding algorithm by
258   training it at each time step and testing it across all time steps. First, the generalization of
259   sensory and action information was compared to chance level across all trials. We
260   observed a statistically significant generalization above chance for action information
261   during the entire trial (p = .005, $\bar{t}_{cluster}$ = 1.36; d = .30) but not for sensory information
262   (**Fig. 3d/f**). These results suggest that priors might differentially impact sensory and action
263   information according to their relative strength.

264   To investigate the effect of prior strength irrespective of their validity. We estimated
265   the common effect of prior strength by summing up the differences in TCP between high
266   and low for both validity conditions (**Fig. 3c**). Strong prior exhibited an enhanced
267   generalization of pre-stimulus sensory information (p = .043, $\bar{t}_{cluster}$ = 1.60; d = .36), but
268   not of action information (**Fig. 3e/g**). Specifically, the decoded sensory information at the
269   end of the cue presentation (p = .03, $\bar{t}_{cluster}$ = -3.75; d = .96) and during the delay period
270   (p = .048, $\bar{t}_{cluster}$ = -2.53; d = .59) reliably generalized across the trial (**Fig. 3i**). These
271   results establish that stronger priors increase the stability of pre-stimulus sensory
272   representation.

273   Together, these results reveal that priors stabilize action and sensory information
274   in distinct ways. Notably, the analyses did not compare how prior validity influences these
275   effects. This raises the question of whether differences in how priors affect sensory and
276   action representations depend on whether the prior information is confirmed.
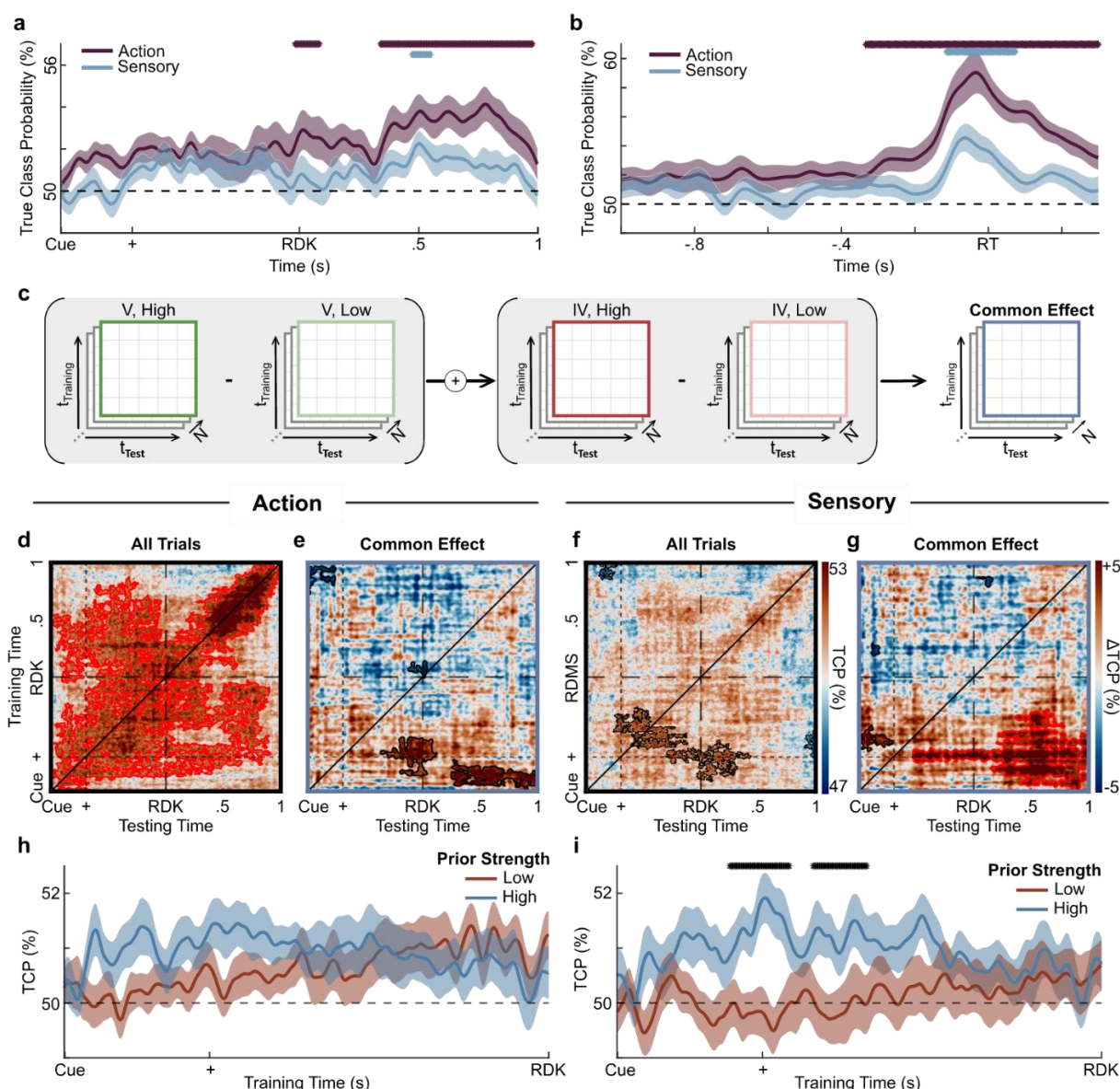
Fig. 3 | **Stronger priors enhance the generalization of pre-stimulus sensory activity. a,** Average true class probability (TCP) of action (left/right button response) or sensory (left or right motion) for each time point across subjects. All decoding analyses were run using a linear discriminant analysis on the data from 257 EEG channels across trials, 10-fold cross-validation, and balancing the trials according to the decoded feature, the prior strength, and cue validity using an oversampling procedure. Data was smoothed with a Gaussian-weighted average over 100 ms window for visualization only. Shaded areas indicate standard deviation. **b,** Same as (**a**) but for data locked to the response. **c,** To estimate the common effect of prior strength irrespective of validity, the low prior condition matrix was subtracted from the high prior condition matrix for both valid (V) and invalid (IV) conditions, and the resulting differences were summed for each subject. **d,** Average generalization performance of action decoding across time for all trials and **f,** sensory decoding. **e,** Estimated common effect of prior strength for each training and test time for action and **g,** sensory decoding. For all time generalization matrices, red contours indicate statistically significant generalization times using a cluster-based permutation test. Black contours indicate the cluster with the highest t-value if none were statistically significant. **h,** Average TCP of action information across all testing times for every pre-stimulus training time. Top lines indicate the time points of significant clusters when testing differences between high and low prior trials using a cluster-based permutation test. **i,** Same as (**h**) but for sensory decoding.

11

Effects of prior validity on neural generalization

On the behavioral level, the effect of strong priors depended on its validity. Individual differences in confirmation bias showed faster and more accurate decisions for valid cues, as well as slower and less accurate decisions for invalid cues (**Fig. 2c**). Additionally, individual differences in speed-accuracy trade-offs showed that faster decisions for valid cues also resulted in impaired accuracy for invalid cues (**Fig. 2d**). It is unclear whether these behavior effects are reflected in the modulation of neural information by (dis)confirmatory evidence. We therefore investigated *whether* and *how* confirmation bias and speed-accuracy trade-offs are correlated with the tuning of neural information by prior validity (**Fig. 4a**).

First, we tested whether the generalization of sensory or action information differs between valid and invalid trials when priors were strong. This method isolates the effects of confirmed versus unconfirmed priors, clarifying how information is tuned over time. Specifically, it determines whether expected sensory and action information is sharpened ($TCP_{Valid,High} > TCP_{Invalid,High}$) or dampened ($TCP_{Invalid,High} > TCP_{Valid,High}$). While sensory information displayed no significant tuning effects, a significant dampening of expected action information was observed (**Fig. 4b/f;** p = .033, $\bar{t}_{cluster}$ = -1.36; d = .30) that was most pronounced for pre-stimulus activity. Furthermore, we compared the mean TCP across all training and test times between valid and invalid trials for each prior strength and decoded feature (**Fig. 4c/g**). Congruent with the time-resolved analysis, there was only a statistically significant validity effect for action information when priors were strong (p = .008, t(19) = 2.98, d = .667). These results suggest that a strong prior induced action dampening, particularly during pre-stimulus activity, while sensory information was not consistently altered across participants.

Neural tuning dynamics is linked to behavior

Having investigated whether and how neural information changes across participants, we tested whether variability in participants' neural tuning were associated with their behavioral performance. Since action was dampened but sensory processing was not (**Fig. 4c/g**), we examined how action dampening, ($TCP_{Invalid,High} - TCP_{Valid,High}$) and

325   sensory sharpening, ($TCP_{Valid,High} - TCP_{Invalid,High}$) correlated with confirmation bias and

326   speed-accuracy trade-offs across participants (**Fig. 4a**).

327   For the confirmation bias, we identified a statistically significant positive correlation

328   cluster between the dampening of pre-stimulus action information before the button

329   response (p = .041, $\bar{r}_{cluster}$ = -.44). In contrast, a negative correlation cluster emerged

330   that extended the generalization of pre-stimulus sensory sharpening to the delay and

331   stimulus presentation (p = .014, $\bar{r}_{cluster}$ = -.48). These findings indicate that participants

332   who dampened pre-stimulus action information before their motor response exhibited

333   greater confirmation bias, while those with stronger sensory sharpening showed less

334   confirmation bias.

335   For the speed-accuracy trade-off (**Fig. 4e/i**), significant clusters were only

336   observed in the correlation with sensory tuning. Specifically, the sharpening of the pre-

337   stimulus sensory information that generalizes to post-stimulus activity was negatively

338   correlated with the speed-accuracy trade-off (p = .043, $\bar{r}_{cluster}$ = -.43), indicating a focus

339   on accuracy. In contrast, the sharpening of post-stimulus sensory information to pre-

340   stimulus activity was positively correlated (p = .023, $\bar{r}_{cluster}$ = .49), suggesting a focus on

341   speed. These findings reveal that speed-accuracy trade-offs are predicted by the

342   dynamics of sensory tuning.

343   Collectively, these results demonstrate that action dampening is linked to stronger

344   reliance on the prior information, while dynamic sensory tuning reflects the balance of
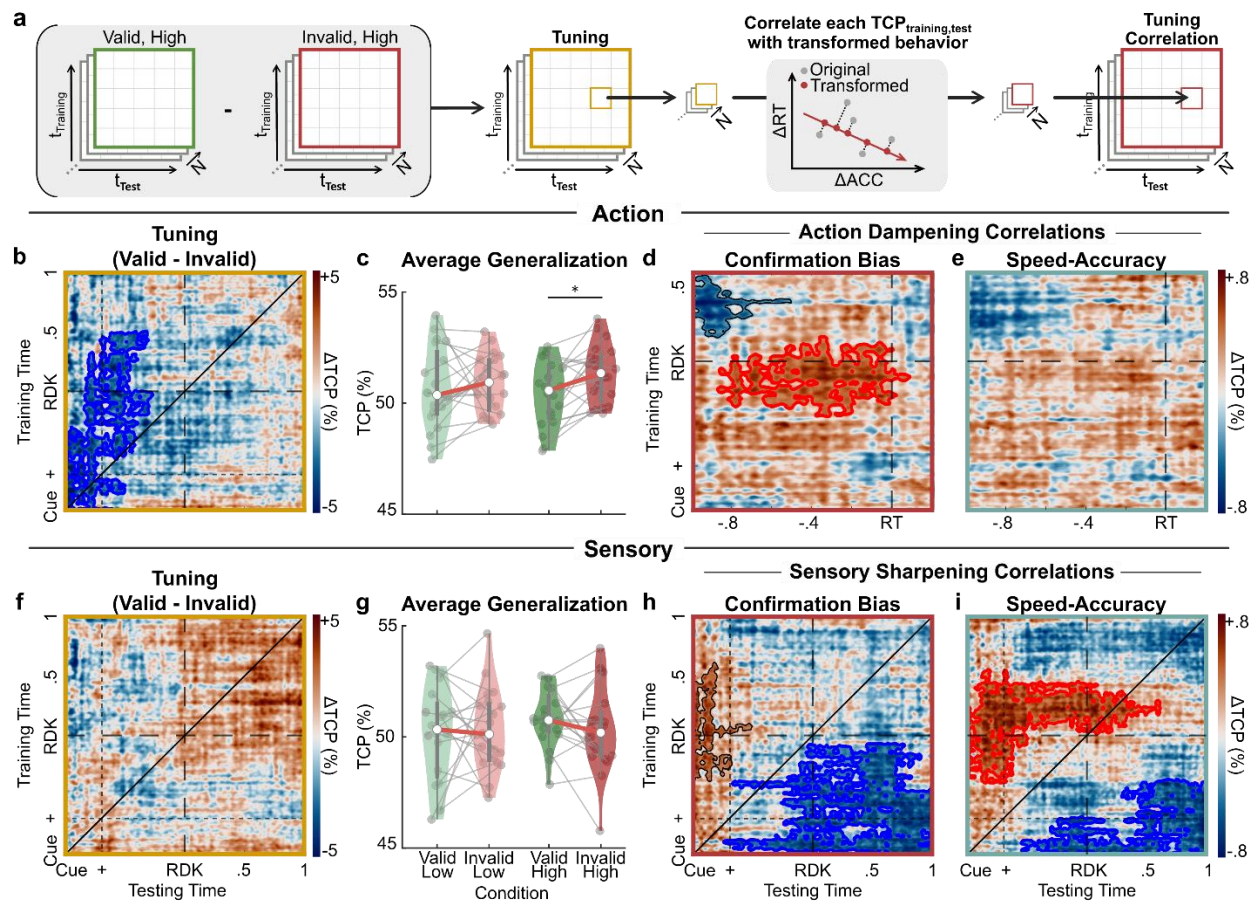
345   costs and benefits of prior integration.

Fig. 4 | Dissociable behavioral correlates of action dampening and sensory sharpening. a, Analysis strategy to trace the temporal evolution of the behavioral effects of action dampening and sensory sharpening. **b,** Grand-mean difference between the cross-temporal generalization of valid and invalid high prior trials ($TCP_{High,Valid} - TCP_{High,Invalid}$) for each training and test time point. **c,** Grand-mean TCP over all training and testing times ([-1, 1] s) for each condition. **d,** Correlation of average cross-temporal action dampening with participant's behavior modulation (transformed behavior from Fig. 2 c) and **e,** with participant's speed-accuracy trade-off (transformed behavior from Fig. 2 d). **f–i,** display the same results as the panels (**b–e**) but for sensory-like activity. Red contours indicate statistically significant generalization times (cluster-based permutation test, p < .05). Black contours highlight non-significant clusters with the highest t-value.

Discussion

Our study reveals how prior expectations influence action and sensory information processes during visual decisions. Behaviorally, prior expectations induced both rigid confirmation bias and flexible speed-accuracy trade-offs, reflecting distinct strategies for handling uncertainty. Neurally, strong priors increased the stability of pre-stimulus sensory information, while action information remained stable regardless of prior strength. By integrating behavioral and neural data, we observed that confirmation biases were

14

365 linked to increased action dampening, while flexibly integrating priors when faced with

366 new evidence was associated with a dynamic sensory tuning process. These findings

367 suggest that prior expectations differentially modulate action and sensory processing,

368 while driving a dynamic sensory tuning that supports flexible decision-making.

369

370 Priors distinctively stabilize sensory and action neural information

371 The central observation of the present study is that neural mechanisms that support the

372 emergence of stable neural representations of sensory and action dynamics findings are

373 distinct. We observed that external prior information (cue information) introduced a stable

374 and long-lasting choice bias, characterized by a stable neural representation as revealed

375 by significant temporal cross-decoding, which emerged already prior to onset of the

376 sensory stimulus. Our results further suggest that this bias was largely independent of

377 cue validity, thus, further contributing to stabilizing the neural representation of the

378 ensuing action.

379 In contrast to action, sensory information was stabilized in a context-dependent

380 manner. We observed that the neural representation of pre-stimulus sensory information

381 was boosted when the prior was strong, but not when the prior was weak. This context-

382 dependent stability of sensory information is well aligned with sensory templates in the

383 early visual cortex[14,15], where confirmatory evidence might further strengthen the

384 preexisting representation. Likewise, it is conceivable that confirmed prior expectations

385 (esp. in states of high confidence) decrease neural variability within[16,17] or across spatial

386 sites[18–20]. In sum, even though representations of prior information are abundant in early

387 sensory as well as motor-related brain areas[21], priors seem to differentially impact

388 sensory and action information processing.

389 By linking neural and behavioral data, our results demonstrate that confirmation

390 biases are linked to increased dampening of action representations as well as reduced

391 sensory sharpening. In contrast, dynamic sensory tuning is associated with, and

392 potentially contributing to, speed-accuracy trade-offs. These observations can be further

393 conceptualized by previous theoretical accounts. For example, cancellation theories

394 suggest a prioritization of unexpected over expected information. This prediction is

395 consistent with the action dampening observed in the current study. Here, action

396 dampening was associated with confirmation biases. Previous studies have also linked
397 confirmation biases to motor preparation[22,23]. These considerations are also fully
398 compatible with a biased starting point of the evidence accumulation process[24,25]. In
399 addition, confirmation biases have also been explained by biased evidence
400 accumulation[26,27]. Computationally, confirmation biases have been shown to benefit
401 hierarchical inference[28] and have been linked to overweighting categorical information
402 during continuous perceptual inference[29]. By linking confirmation biases with neural
403 information tuning, the present study offers a possible mechanistic explanation that
404 connects evidence accumulation studies to perceptual prediction theories.

405   In addition to cancellation theories, the Opposing Process Theory[8] proposes that
406 sensory information is initially sharpened by prior expectations that can be subsequently
407 dampened if bottom-up processing signals conflict with prior expectations. In line with this
408 theory, we observed that sharpening of pre-stimulus sensory information predicted faster
409 decisions when the prior was confirmed but at the cost of less accurate decisions when
410 the prior was disconfirmed. In contrast, dampening sensory information during the
411 stimulus presentation benefitted task accuracy by overriding the prior when it was
412 disconfirmed but resulted in slower reaction times. Hence, our present results also
413 provide supporting evidence for the Opposing Process Theory in balancing speed and
414 accuracy during visual decisions. In sum, the present results reconcile and dissociate
415 several previous theoretical and empirical findings and indicate that neural
416 implementations of priors differentially impact sensory and motor processes. These
417 considerations raise the question of how different perceptual prediction theories can be
418 reconciled or co-exist.

419   A central question—and a key conflict in existing literature—is whether sensory
420 processes sharpen or dampen expected information[6,8]. Our results are incompatible with
421 one central prior coding scheme and rather suggest that independent coding schemes
422 exist for the sensory and action domains. Critically, our behavioral results confirm that
423 each scheme has a differential impact on behavioral performance. These findings raise
424 an important question: Are the differences between action and sensory domain inherent,
425 or are they shaped by the task demands?

426

Inherit or functional differences between sensory and action domains

Many accounts of perceptual decision-making characterize sensory and action coding as parallel processes, where sensory evidence accumulation and motor preparation occur simultaneously[30]. Under this view, sensory evidence continuously flows from sensory to motor areas. While several lines of inquiry provided convincing evidence supporting concurrent action integration in motor areas of humans[31] as well as monkeys[32,33], recent findings challenged this view. For example, intracranial recordings in patients with epilepsy revealed effector-independent evidence accumulation signals across diverse perceptual decision tasks[34]. In addition, perceptual decisions seem to be formed before action preparation when evidence must be fully integrated before forming a motor plan[35]. Complementary results from large-scale neural recordings in mice further demonstrate that sensory integration in premotor neural populations operates in a parallel fashion but transitions sharply to an orthogonal action execution subspace once evidence accumulation is completed[36]. These findings across studies and species indicate that sensory evidence accumulation and action execution are inherently distinct processes during perceptual decision-making.

While sensory and motor processing are widely supported to occur in relative isolation, the distinctive impact of prior information in these domains might be functional. Under a Bayesian Decision Theory perspective, the posterior belief (i.e., the final choice) depends on the prior information and its likelihood (i.e., the current sensory evidence). Under this view, both sensory evidence and prior information influence the perceptual choice, and thereby, the ensuing action execution. However, as in most experimental designs, only the sensory input can provide immediate additional evidence to update the posterior belief. It is possible that a dynamic neural tuning of predictive information only occurs when it is decision-relevant. Congruent with this idea, prior-induced sensory modulations were only observed when the feature was goal-directed[37,38]. Similarly, dampening of predictive information may occur when the information is not needed to update the posterior, such as with action information in most perceptual tasks. These considerations explain why cancellation theories are prevalent in the action and motor literature.

17

Limitations

Our study raises considerations that can be explored in future research. First, to experimentally distinguish the formation of a choice variable from the action execution, future studies could exploit experimental designs that prevent the formation of action plans with variable sensory-to-response mappings, such as in Sandhaeger et al.[39]. Second, participants had to choose between two cues, introducing an additional level of uncertainty, particularly due to the inherent exploration-exploitation trade-off when learning under stochastic regimes[40]. The influence of priors during exploratory behavior may differ from their effects during exploitation, enhancing variability in how priors affect sensory tuning. Finally, a sensory sharpening was not consistently observed across participants, raising questions about the timescales of priors and their impact on perceptual processes[5]. Shorter timescales may introduce greater individual variability, while longer timescales yield more homogeneous results. In support of this hypothesis, sensory sharpening has been reported only after extended task exposure[41], with expectation suppression similarly observed following prolonged training[42]. Additionally, the consolidation of prior information during sleep may account for persistent predictions in early visual areas[43]. Longer-term priors, such as the statistics of natural environment, may optimize sensory processing through mechanisms consistent with efficient coding theories[6,44]. In contrast, volatile environments or tasks with shorter timescales require learning a hierarchical structure, which may increase individual variability.

Conclusion

This study demonstrates that priors can be integrated rigidly or flexibly. Confirmation biases (a rigid tendency to rely on cues) was linked to increased dampening of action information and reduced sensory sharpening. In contrast, speed-accuracy trade-offs (i.e., more flexible prior influence) were attributed to a dynamic sensory tuning mechanism, adjusted based on whether sensory evidence was confirmatory. While sustained sharpening of sensory information predicted faster responses when the prior was confirmed, dampening of expected sensory information upon disconfirmatory evidence overrode the prior; thus, leading to correct decisions. These findings support recent

18

488 theories of predictive processing and suggest that dynamically tuning prior sensory

489 information according to current evidence is essential for flexible decision-making.

490

Method

Participants

20 healthy volunteers (10 females, 20 to 31 years) participated in this study and were recruited from the student community. All participants had normal vision and no photosensitive epilepsy, claustrophobia, color blindness, or other neurological disorders. 19 participants were right-handed. Participants provided informed written consent to participate in the study and were paid for their participation according to their performance. This study was approved by the Ethics Committee of the Medical Faculty Tübingen (protocol number 049/2020B02) and conducted in accordance with the Declaration of Helsinki.

Experimental Design

Before the main task, the motion coherence threshold was estimated through QuestPlus, an adaptive psychometric method[45,46] to yield 75% accuracy. The motion stimuli consisted of a sequence of moving white dots (size: .06 visual angle; density: 16.7 dots/degree; velocity: .03 visual angle/second; lifetime: 12 frames) inside a circular area, centered on a fixation cross, with a diameter of 2.5 visual angle degrees, on a black background. Most dots moved in random directions, while a small subset (up to 20%) coherently moved either to the left or right. Participants were seated 120 cm from a 28" LED screen (60 Hz, 2560x1440 px) in a dark, magnetically and acoustically shielded room. Participants were instructed to respond by pressing the buttons on two response boxes (one for each hand). All stimuli were presented using PsychToolbox for MATLAB[47,48].

Behavioral Task

The behavioral task combines reversal learning and a random dot motion discrimination task. Each trial contained three main decisions: cue choice, confidence, and motion

518   direction, respectively. The main trial events were interleaved with fixation crosses, as

519   shown in **Fig. 1a**. The start of the trial was self-paced. The cue identity was established

520   by distinct colors (blue or orange). Immediately after the cue choice, participants rated

521   their confidence that it would be valid. The cue direction was displayed for 300 ms, and

522   its validity was probabilistically determined according to an underlying pseudo-

523   randomized structure (see true contingency in **Fig. 1c**). Participants were familiarized with

524   the main task with 15 training trials, in which no validity reversals occurred.

525   Cues within informative blocks had validities of 80% and 30%. The best cue was

526   pseudo-randomized across informative blocks, such that the same cue was never the

527   best on three consecutive informative blocks. Non-informative blocks had both cue

528   validities set at 30%. Cue validity for single trials was pseudo-randomized to have an

529   exponential moving average of the past 5 trials around the target probability (±15%). An

530   informative block lasted between 15 and 30 trials. Switching from an informative block to

531   a non-informative block occurred before 30 trials if a successful learning criterion was

532   matched (exponential moving average of the accuracy for the past 10 cue choices above

533   80%) after the 15th trial in the block. Each non-informative block lasted 15 trials.

534   Participants completed a maximum of 320 trials.

535

536   **Reward Structure**

537   At the end of each trial, participants received two rewards: one for the motion judgment

538   and another for the cue choice. For correct motion responses, 10 points were assigned.

539   For incorrect motion responses, 10 points were deducted. The reward for the cue choice

540   depended on the confidence reported earlier in the trial. If participants did not respond

541   within the time limit, a negative reward was assigned (-20 points), and the trial was

542   terminated. The reward of the cue choice was dependent on the actual validity of the cue

543   and the confidence rated by the participant on a particular trial. Reward values were

544   estimated based on a proper scoring rule[49], in which participants received 10 points if

545   their confidence was high or 8 points if their confidence was low, while they lost 9 points

546   if their confidence was high or 4 points if their confidence was low.

20

Behavioral Modeling

The latent prior strength of the optimal cue across trials was modeled using a variant of the Kalman Filter[50], the Volatile Kalman Filter[12] (VKF). In brief, its generative model assumes that the two hidden Markov chains, in which the latent prior indicates the optimal cue for a given trial, x_t, diffuses over time according to the hidden (inverse) volatility of the environment, z_t, such that $x_t = N(x_t|x_{t-1}, z_t^{-1})$. In its turn, z_t is given by its previous values and evolves dynamically according to a multiplicative diffusion noise, $z_t = z_{t-1}\epsilon_t(\lambda)$. ϵ_t is defined as a rescaled beta distribution such that $\epsilon_t \in [0, (1-\lambda)^{-1}]$, and $0 < \lambda < 1$. Therefore, the free parameter λ controls the diffusion of the volatility. On the first trial, the inverse mean of z_t is given by a free parameter v_0. The choice, y_t, is defined as a softmax function of the difference between the latent prior for each cue with a decision exploration (free) parameter β:

$$y_t = \frac{1}{1 + \exp\left(-\beta\left(x_t - (1 - x_t)\right)\right)}.$$

The outcome of a given trial, $o_t \in \{0,1\}$, is given by $o_t = \text{Bernoulli}\left(\sigma(x_t)\right)$, where $\sigma(x_t) = 1/1 + e^{-x_t}$. In our case, we assume that the cue validity of each trial provides evidence for a mutually exclusive state of whether blue is currently the best cue such that

$$o_t = \begin{cases} 0, & \text{if the chosen cue was blue (invalid) or orange (valid)}, \\ 1, & \text{if the chosen cue was blue (valid) or orange (invalid)}. \end{cases}$$

For the subsequent analysis, we used the prior of the chosen cue being the best, $\text{prior} = x_t \cdot \mathbb{I}(\text{chosen cue = Blue}) + (1 - x_t) \cdot \mathbb{I}(\text{chosen cue = Orange})$, where $\mathbb{I}(\cdot)$ equals 1 if the condition inside is true and 0 otherwise. The prior was then categorized into low and high using a subject-level median split.

Model Fitting and Comparison

To evaluate the performance of the VKF, we consider three alternative learning models: the Rescorla-Wagner[51] (RW), the Kalman Filter[50] (KF), and the Hierarchical Gaussian Filter[52,53] (HGF). While the KF and the HGF models account for volatilities in the environment, the RW model does not. A softmax function with a free temperature parameter, β, was used as a response model for all learning models as described for the VKF.

576　　　　The models were fit and compared (*Supplementary Figure 3*) with a hierarchical
577　Bayesian inference approach based on variational Bayes[54], in which fitted parameters
578　are assumed to be drawn from a normal distribution based on the group-level statistics
579　and the model identity is considered a random effect. To estimate the subject-level
580　posterior, a Laplace approximation was used.

581

582　**EEG Data Acquisition and Preprocessing**
583　Electrophysiological (EEG) data were continuously recorded with 257 scalp electrodes
584　with a 1000 Hz sampling rate and an online reference at Cz (Electrical Geodesics, Inc.).
585　The data were preprocessed using the FieldTrip[55] toolbox for Matlab by applying a
586　bandpass Butterworth filter (0.3 to 40 Hz), demeaning, and re-referencing to the common
587　average. Trials with muscle artifacts were visually identified and excluded. Eye and
588　muscle artifacts were manually removed using Independent Component Analysis. The
589　data were down-sampled to 500 Hz, epoched from -1000 to 1000 ms relative to the RDK
590　onset, and z-scored using the trial-wise mean and standard deviation from the -400 to 0
591　ms interval relative to the cue presentation.

592

593　**Univariate Analysis**
594　Time-locked responses were calculated per participant and electrode by averaging
595　across different experimental conditions relative to the onset of the motion stimulus or
596　relative to response onset. Differences between event-related potentials (ERPs) were
597　assessed using cluster-based permutation statistics as implemented in FieldTrip[55].

598

599　**Multivariate Analysis**
600　Decoding of action (left or right button press) or sensory-like activity (left or right motion)
601　was performed on 257 channels using a Linear Discriminant Analysis[13,56] using the
602　MVPA-Light[57] toolbox for Matlab. In all decoding analyses, we performed a 10-fold cross-
603　validation procedure balancing the selected feature (action or sensory), the prior strength
604　(low or high), and the cue validity (valid or invalid) across the training samples, but not
605　the testing samples, in each fold by upsampling. The true class probability (TCP) of testing
606　samples was estimated based on the multivariate distribution for the training fold.

607   To compare the evolution of decoded signals, we performed a time generalization

608   analysis across all time points and trials, resulting in a $N_{trials} \times N_{samples} \times N_{samples}$ TCP

609   matrix per participant. Then, this matrix was averaged across conditions, resulting in a

610    $TCP_{prior,validity}$ matrix with $N_{samples} \times N_{samples}$ per participant. For each decoded feature,

611   the common effect of having a high prior (**Fig. 3 c**) is defined by

612

613   Common effect $= \left[\text{TCP}_{\text{High, Valid}} - \text{TCP}_{\text{Invalid, Low}}\right] + \left[\text{TCP}_{\text{Invalid, High}} - \text{TCP}_{\text{Invalid, Low}}\right],$

614

615   isolating the effect of having a high prior regardless of the validity of the cue.

616   The tuning (sharpening if positive, dampening if negative) of the decoded signals

617   was defined as the difference between the valid and invalid conditions within high prior

618   trials (**Fig. 4 a**),

619

620   Tuning $= \left[\text{TCP}_{\text{High, Valid}} - \text{TCP}_{\text{High, Invalid}}\right].$

621

622   The sensory sharpening and action dampening for each training and test time were

623   then correlated with the behavioral modulation and speed-accuracy trade-offs. To capture

624   the individual differences (**Fig. 1 c/d**), the participant's values of the two correlated

625   features were projected into their regression line, capturing their shared variance along a

626   single dimension. This procedure was repeated for response-locked $TCP_{prior,validity}$ data.

627

628   **Statistical Analysis**

629   Unless stated otherwise, we employed paired t-tests. All electrophysiological recordings

630   were assessed using cluster-based permutation tests as implemented in FieldTrip[55] that

631   were based on paired t-tests. Clusters were determined across space and time (univariate

632   analyses) or across time (multivariate analyses) by summing t-values within the

633   thresholded cluster. The observed cluster was compared to a surrogate distribution

634   (10.000 shuffles) where condition labels were randomly permuted between groups. Effect

635   sizes were quantified using Cohen's d.

636

Data and code availability

All data and custom code used for analyses are available from the corresponding author upon request. Preprocessing of EEG data and cluster-based permutation tests were performed using Fieldtrip[55] toolbox for Matlab and decoding analysis where performed using the MVPA-Light[57] toolbox for Matlab.

References

1. Fritsche, M. *et al.* Temporal regularities shape perceptual decisions and striatal dopamine signals. *Nat Commun* **15**, 7093 (2024).

2. Urai, A. E., de Gee, J. W., Tsetsos, K. & Donner, T. H. Choice history biases subsequent evidence accumulation. *eLife* **8**, e46331 (2019).

3. Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M. & Donner, T. H. Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence. *Current Biology* **28**, 3128-3135.e8 (2018).

4. Knill, D. C. & Richards, W. *Perception as Bayesian Inference.* (Cambridge University Press, 1996).

5. Gold, J. I. & Stocker, A. A. Visual Decision-Making in an Uncertain and Dynamic World. *Annual Review of Vision Science* **3**, 227–250 (2017).

6. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends in Cognitive Sciences* **22**, 764–779 (2018).

7. Blakemore, S. J., Wolpert, D. M. & Frith, C. D. Central cancellation of self-produced tickle sensation. *Nat Neurosci* **1**, 635–640 (1998).

8. Press, C., Kok, P. & Yon, D. The Perceptual Prediction Paradox. *Trends in Cognitive Sciences* **24**, 13–24 (2020).

9. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* **75**, 265–270 (2012).

10. Summerfield, C. & Koechlin, E. A Neural Representation of Prior Information during Perceptual Inference. *Neuron* **59**, 336–347 (2008).

11. Kilteni, K. & Ehrsson, H. H. Body ownership determines the attenuation of self-generated tactile sensations. *Proceedings of the National Academy of Sciences* **114**, 8426–8431 (2017).

12. Piray, P. & Daw, N. D. A simple model for learning in volatile environments. *PLOS Computational Biology* **16**, e1007963 (2020).

13. Fisher, R. A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* **7**, 179–188 (1936).

14. Kok, P., Failing, M. F. & de Lange, F. P. Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience* **26**, 1546–1554 (2014).

15. Kok, P., Mostert, P. & de Lange, F. P. Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences* **114**, 10473–10478 (2017).

16. Geurts, L. S., Cooke, J. R. H., van Bergen, R. S. & Jehee, J. F. M. Subjective confidence reflects representation of Bayesian probability in cortex. *Nat Hum Behav* (2022) doi:10.1038/s41562-021-01247-w.

17. Skowron, A. *et al.* Neural variability compresses with increasing belief precision during Bayesian inference. 2024.01.11.575180 Preprint at https://doi.org/10.1101/2024.01.11.575180 (2024).

18. van Bergen, R. S. & Jehee, J. F. M. Modeling correlated noise is necessary to decode uncertainty. *NeuroImage* **180**, 78–87 (2018).

19. Lange, R. D. & Haefner, R. M. Task-induced neural covariability as a signature of approximate Bayesian learning and inference. *PLOS Computational Biology* **18**, e1009557 (2022).

20. van den Brink, R. L. *et al.* Flexible sensory-motor mapping rules manifest in correlated variability of stimulus and action codes across the brain. *Neuron* **111**, 571-584.e9 (2023).

21. Findling, C. *et al.* Brain-wide representations of prior information in mouse decision-making. 2023.07.04.547684 Preprint at https://doi.org/10.1101/2023.07.04.547684 (2023).

22. de Lange, F. P., Rahnev, D. A., Donner, T. H. & Lau, H. Prestimulus Oscillatory Activity over Motor Cortex Reflects Perceptual Expectations. *Journal of Neuroscience* **33**, 1400–1410 (2013).

23. Kelly, S. P., Corbett, E. A. & O'Connell, R. G. Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nat Hum Behav* **5**, 467–481 (2021).

24. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review* **113**, 700–765 (2006).

25. Rao, V., DeAngelis, G. C. & Snyder, L. H. Neural Correlates of Prior Expectations of Motion in the Lateral Intraparietal and Middle Temporal Areas. *J. Neurosci.* **32**, 10063–10074 (2012).

26. Rollwage, M. *et al.* Confidence drives a neural confirmation bias. *Nat Commun* **11**, 2634 (2020).

27. Park, H. *et al.* Confirmation Bias through Selective Use of Evidence in Human Cortex. 2024.06.21.600060 Preprint at https://doi.org/10.1101/2024.06.21.600060 (2024).

28. Qiu, C., Luu, L. & Stocker, A. A. Benefits of commitment in hierarchical inference. *Psychological Review* **127**, 622–639 (2020).

29. Lange, R. D., Chattoraj, A., Beck, J. M., Yates, J. L. & Haefner, R. M. A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLOS Computational Biology* **17**, e1009517 (2021).

30. Donner, T. H., Siegel, M., Fries, P. & Engel, A. K. Buildup of Choice-Predictive Activity in Human Motor Cortex during Perceptual Decision Making. *Current Biology* **19**, 1581–1585 (2009).

31. Wilming, N., Murphy, P. R., Meyniel, F. & Donner, T. H. Large-scale dynamics of perceptual decision information across human cortex. *Nat Commun* **11**, 5109 (2020).

32. Roitman, J. D. & Shadlen, M. N. Response of Neurons in the Lateral Intraparietal Area during a Combined Visual Discrimination Reaction Time Task. *J. Neurosci.* **22**, 9475–9489 (2002).

33. Shushruth, S., Mazurek, M. & Shadlen, M. N. Comparison of Decision-Related Signals in Sensory and Motor Preparatory Responses of Neurons in Area LIP. *J. Neurosci.* **38**, 6350–6365 (2018).

34. Gherman, S. *et al.* Intracranial electroencephalography reveals effector-independent evidence accumulation dynamics in multiple human brain regions. *Nat Hum Behav* **8**, 758–770 (2024).

731   35. Lisi, M., Morgan, M. J. & Solomon, J. A. Perceptual decisions and oculomotor
732   responses rely on temporally distinct streams of evidence. *Commun Biol* **5**, 189
733   (2022).

734   36. Khilkevich, A. *et al.* Brain-wide dynamics linking sensation to action during decision-
735   making. *Nature* **634**, 890–900 (2024).

736   37. Park, J., Kim, S., Kim, H. R. & Lee, J. Prior expectation enhances sensorimotor
737   behavior by modulating population tuning and subspace activity in sensory
738   cortex. *Science Advances* **9**, eadg4156 (2023).

739   38. Ruesseler, M., Weber, L. A., Marshall, T. R., O'Reilly, J. & Hunt, L. T. Quantifying
740   decision-making in dynamic, continuously evolving environments. *eLife* **12**,
741   e82823 (2023).

742   39. Sandhaeger, F., Omejc, N., Pape, A.-A. & Siegel, M. Abstract perceptual choice
743   signals during action-linked decisions in the human brain. *PLOS Biology* **21**,
744   e3002324 (2023).

745   40. Domenech, P., Rheims, S. & Koechlin, E. Neural mechanisms resolving
746   exploitation-exploration dilemmas in the medial prefrontal cortex. *Science* **369**,
747   eabb0184 (2020).

748   41. Walsh, K., McGovern, D. P., Dully, J., Kelly, S. P. & O'Connell, R. G. Prior
749   probability cues bias sensory encoding with increasing task exposure. *eLife* **12**,
750   RP91135 (2024).

751   42. Feuerriegel, D., Vogels, R. & Kovács, G. Evaluating the evidence for expectation
752   suppression in the visual system. *Neuroscience & Biobehavioral Reviews* **126**,
753   368–381 (2021).

754   43. Yon, D. *et al.* Stubborn Predictions in Primary Visual Cortex. *Journal of Cognitive*
755   *Neuroscience* **35**, 1133–1143 (2023).

756   44. Simoncelli, E. P. & Olshausen, B. A. Natural Image Statistics and Neural
757   Representation. *Annual Review of Neuroscience* **24**, 1193–1216 (2001).

758   45. Jones, P. R. QuestPlus: A MATLAB Implementation of the QUEST+ adaptive
759   Psychometric Method. *Journal of Open Research Software* **6**, 27 (2018).

760   46. Watson, A. B. QUEST+: A general multidimensional Bayesian adaptive
761   psychometric method. *Journal of Vision* **17**, 10 (2017).

762   47. Brainard, D. H. The Psychophysics Toolbox. *Spatial Vis* **10**, 433–436 (1997).

763   48. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming

764   numbers into movies. *Spatial Vis* **10**, 437–442 (1997).

765   49. Carvalho, A., Dimitrov, S. & Larson, K. On proper scoring rules and cumulative

766   prospect theory. *EURO J Decis Process* **6**, 343–376 (2018).

767   50. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Journal

768   of Basic Engineering* **82**, 35–45 (1960).

769   51. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning : Variations in

770   the effectiveness of reinforcement and non-reinforcement. in *Classical

771   conditioning II: Current research and theory* (eds. Black, A. H. & Prokosy, W. F.)

772   vol. 2 64–69 (Appleton-Century-Crofts, New Yorker, 1972).

773   52. Mathys, C. D. *et al.* Uncertainty in perception and the Hierarchical Gaussian Filter.

774   *Front. Hum. Neurosci.* **8**, (2014).

775   53. Mathys, C., Daunizeau, J., Friston, K. & Stephan, K. A Bayesian Foundation for

776   Individual Learning Under Uncertainty. *Frontiers in Human Neuroscience* **5**,

777   (2011).

778   54. Piray, P., Dezfouli, A., Heskes, T., Frank, M. J. & Daw, N. D. Hierarchical Bayesian

779   inference for concurrent model fitting and comparison for group studies. *PLOS

780   Computational Biology* **15**, e1007043 (2019).

781   55. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source

782   Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological

783   Data. *Computational Intelligence and Neuroscience* **2011**, e156869 (2010).

784   56. Zhao, S., Zhang, B., Yang, J., Zhou, J. & Xu, Y. Linear discriminant analysis. *Nat

785   Rev Methods Primers* **4**, 1–16 (2024).

786   57. Treder, M. S. MVPA-Light: A Classification and Regression Toolbox for Multi-

787   Dimensional Data. *Front. Neurosci.* **14**, (2020).

788