An integral role for timing in interception

Chia-Jung Chang[1] and Mehrdad Jazayeri[1,2]

[1]Department of Brain and Cognitive Science, [2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Correspondance should be addressed to M.J. (mjaz@mit.edu)
Mehrdad Jazayeri, PhD
Robert A. Swanson Career Development Professor
Assistant Professor, Dept. of Brain and Cognitive Sciences
Investigator, McGovern Institute for Brain Research
43 Vassar Street, Building 46-6041,Cambridge, MA 02139–4307
Massachusetts Institute of Technology
Phone: 617–715–5418
Email: mjaz@mit.edu

Author Contribution

C.-J.C. designed the experiment, collected and analyzed the data, and performed the computational modeling. M.J. helped in experimental design, data analysis, and provided intellectual support throuout the study. C.-J.C. and M.J. wrote the manuscript.

Abstract

1   Timing is critical for myriad behaviors in dynamic environments. For example, to intercept an object, the

2   brain must compute a reliable estimate of time-to-contact (TTC). Prior work suggests that humans

3   compute TTC using kinematic information such as distance and speed without explicitly relying on

4   temporal cues, just as one would do in a physics classroom using kinematic equations. Considering the

5   inherent uncertainty associated with estimates of speed and distance and the ability of human brain to

6   combine different sources of information, we asked whether humans additionally rely on temporal cues.

7   We found that humans actively integrate speed information with both explicit and implicit timing cues.

8   Analysis of behavior in relation to a Bayesian model revealed that the additional temporal information

9   helps subjects optimize their performance in the presence of measurement uncertainty. These findings

10   suggest that brain's timing mechanisms are actively engaged while interacting with dynamic stimuli.

2

Introduction

11 Imagine intercepting a moving ball on a pool table as it bounces around hitting different edges. Does one

12 only process kinematic information such as distance, position and speed, or does one additionally pay

13 attention to *when* the ball hits the edges? At first glance, the answer seems trivial: since kinematic

14 variables and time are directly related through kinematic equations (e.g., t = d/v), there would be no

15 additional advantage in tracking time. What if internal estimates of speed and position were unreliable,

16 for example, the lights were too dim to clearly see the ball? In that case, one may choose to pay attention

17 to when the ball hits the edges to improve accuracy. This example highlights a general, important, and

18 unresolved question in sensorimotor processing: do humans actively engage brain's timing mechanisms

19 when interacting with dynamic stimuli, or do they solely rely on kinematic information?

20 We aimed to address this question using a virtual object interception task. When intercepting a moving

21 object, one has to estimate when the object reaches a desired target location, a variable that we will refer

22 to as time-to-contact (TTC). Early studies hypothesized that TTC is derived from the rate of expansion of

23 an object's retinal image[1–3]. Later, it was suggested that TTC is derived indirectly from an object's speed

24 and position[4–13]. For example, when an object moves with a fixed speed, TTC would be computed by

25 dividing perceived distance by perceived speed. However, as the example of the pool table illustrates, the

26 inherent variability in the measurement and processing of kinematic information[14–20] renders temporal

27 cues highly relevant for the estimation of TTC. Humans' ability to measure time intervals independent of

28 kinematic cues is well-documented[21–24]. Furthermore, decades of research indicate that humans can

29 efficiently combine multiple sources of information[14, 25–32]. Therefore, we hypothesized that humans

30 integrate kinematic information with explicit and implicit temporal cues to derive better estimates of TTC

31 (**Box 1**).

32   To test this hypothesis, we designed a series of experiments in which the subjects had to press a key when

33   a bar moving along a linear path would arrive at a target position (**Fig. 1a**). While moving, the bar was

34   sometimes visible and sometimes occluded. By varying the temporal structure between visible and

35   occluded part of the path, we manipulated the reliability of the speed and the temporal information

36   independently to compute TTC. Consistent with our hypothesis, we found that subjects actively used

37   temporal information to improve their estimates of TTC.

38   To better understand the nature of the underlying computations, we compared subjects' behavior to that of

39   an ideal Bayesian observer who optimally integrates speed and timing information. Similar to work in

40   other sensorimotor domains[14, 25–32], the model was able to accurately capture subjects' estimation strategy

41   indicating that humans efficiently integrate prior statistics with measurements of both speed and elapsed

42   time. These results highlight a hitherto unappreciated function of the brain's capacity to utilize time –

43   independent of speed and distance – to inform sensorimotor function while interacting with dynamic

44   stimuli.

Results

Experiment 1: Interception performance benefits from explicit timing cues

45   We first asked whether humans are capable of integrating speed information with temporal cues to

46   improve their estimate of TTC in an interception task. To do so, we asked subjects to intercept a moving

47   bar in three conditions: one with a speed cue, one with a timing cue, and one with both cues present (**Fig.

48   2a**). In the first condition, the bar was visible only in the early part of the path, and then was occluded

49   before it reached the target position. Subjects had to measure the speed of the bar from an early visible

50   segment and use that to compute when the bar would reach the target position at the end of the occluded

51   segment. We denote this condition by I(VM) as a shorthand for Interception in the presence of Visible

52   Motion.

4

53 In the second condition, the motion was invisible but the position of the bar was flashed at the beginning

54 of the path and when it reached the central fixation point, which was halfway along the path. TTC had to

55 be computed based on the interval between two flashes along the path. We placed the timing cue around

56 the fixation point to avoid causing a gaze shift in response to the flash, which could impact the subjects'

57 overall estimation strategy. We denote this condition by I(FP) as a shorthand for Interception in the

58 presence of Flashed Position.

59 In the third condition, the bar was visible early on, and additionally, its position was flashed briefly at

60 central fixation point, i.e., halfway between the initial and target location, giving subjects the opportunity

61 to measure both speed and timing information. We denote this condition by I(VM+FP).

62 The actual time-to-contact (*TTCa*) was defined as the time from when the bar passed the central fixation

63 to when it reached the target. We compared *TTCa* to the interval between when the bar passed the central

64 fixation and subjects pressed the button. We refer to this interval as the produced time-to-contact (*TTCp*).

65 For I(FP) and I(VM+FP) conditions, *TTCp* was straightforwardly computed from the time when the bar

66 was flashed to when the button was pressed. For the I(VM) condition, because the bar was not visible

67 throughout the path (not flashed at the fixation point), we estimated *TTCp* by appropriately scaling the

68 response time by the occluded segment.

69 As evident from the *TTCp* pattern for a typical subject (**Fig. 2a**), subjects were able to perform the task in

70 all three conditions with different degrees of of sensitivity. *TTCp* values were variable and systematically

71 biased toward the mean. We quantified this regression to the mean by computing a BIAS term that

72 quantifies the overall deviation from the identity line (see Methods; **Fig. 2b**). The BIAS was significantly

73 smaller when both cues were available compared to the speed cue condition ($t_{198} = 26.6435$, $p < 0.001$,

74 hedges' $g = 3.7537$) as well to the timing cue condition ($t_{198} = 27.4602$, $p < 0.001$, hedges' $g = 3.8687$).

75 This reduction in BIAS was observed for all the subjects (**Fig. 2c**) and was significant across subjects

76   (Wilcoxon one-side signed-rank test, statistics = 28, p < 0.01), suggesting that humans are capable of

77   integrating speed information with temporal cues to reduce uncertainty.

78   While this result is consistent with subjects integrating the two cues, it is also possible that the subjects

79   did not integrate the two cues and instead used the timing cue (flash at the fixation point) to simply reset

80   their subjective estimate of the position of the bar to the central fixation. To test this possibility, we tested

81   a subset of subjects in a cue conflict paradigm in which the flash at the central fixation (FP) was jittered

82   by –100, 0, or 100 msec relative to when the bar reached the central fixation (**Supplementary Fig. 1a**).

83   To understand the logic of this experiment, let us predict subjects' estimated time-to-contact (*TTCe*)

84   under various hypotheses for a case where the flash is presented 100 ms later that when the bar reaches

85   the fixation point. If a subject only relies on the timing cue, they would overestimate the time from

86   motion onset to the flash by 100 msec; this would predict that *TTCe* would be 100 msec longer than

87   *TTCa*. If a subject only relies on speed, then the lagging flash would not influence the subject's behavior.

88   However, since we quantify *TTCp* with respect to the time when the bar is flashed, we would register a

89   *TTCe* that would be 100 shorter than *TTCa*. As a third hypothesis, let us consider that a subject would use

90   the flash to reset the position, and use the speed to estimate *TTCe*. For this hypothesis, the *TTCe* would

91   remain the same as when the subject only used the speed cue. Finally, if the subjects were to integrate the

92   two cues, we would see a bias in *TTCe* in the direction of the jitter that would be less than 100 msec away

93   from the non-jittered condition. The results were consistent with the last hypothesis of integration and

94   could not be explained by hypotheses including the position-reset hypothesis (**Supplementary Fig. 1a**).

Experiment 2: Interception performance benefits from inherent timing cues

95   Experiment 1 demonstrated that humans were able to integrate timing cue with speed information.

96   However, this could have been due to the presentation of an explicit timing cue (i.e., the flash a the

97   fixation point). Therefore, we asked whether subjects utilize the timing information from the visible

98   portion of the motion even if no explicit flash at the fixation point is presented. To validate the role of

6

99   time as an additional cue, it was important to make sure that longer visible motion did not additionally

100 improve subjects' estimate of the speed. Therefore, as a first step, we measured the ability of subjects to

101 estimate speed during interception with different visible lengths ranging from 0.625 to 5 degree in log

102 scale while keeping the occluded distance fixed (**Supplementary Fig. 2**). We evaluated performance by

103 measuring subjects' root mean squared error (RMSE). We found that performance improved significantly

104 as the visible lengths increased from 0.625 to 1.25 degree (paired-sample t-test, $t_{399} = 56.61$, $p < 0.001$)

105 and saturated afterwards (paired-sample t-test, $t_{399} = 0.9031$, $p = 0.3670$). In other words, the fidelity of

106 the speed estimate saturated at a visible length of 1.25 degree.

107 We then conducted an interception task where we evaluated the relevance of the timing cue by changing

108 the visible length compared to the occluded length. Importantly, in all conditions, the visible length was

109 beyond the saturation point in the I(VM) task. This ensured that any improvement in performance was not

110 due to an improvement of speed estimates. We tested subjects' performance in three conditions. In all

111 conditions, the occluded length was fixed ($d_2 = 8$ degree). Across conditions, the ratio of the occluded

112 length (d_2) to the visible length (d_1) was varied by a gain factor ($G = d_2/d_1$). The three gain factors were

113 0.667, 1, and 1.6.

114 **Figure 3a and 3b** shows the performance of a typical subject in the three conditions. Surprisingly, the

115 best performance was not associated with $G = 0.667$ when the visible length was longest. Instead, RMSE

116 was smallest when the visible and occluded lengths were equal ($G = 0.667$, $t_{198} = 20.3981$, $p < 0.001$,

117 hedges' $g = 2.9308$; $G = 1.6$, $t_{198} = 22.9261$, $p < 0.001$, hedges' $g = 3.2299$), which we refer to as the

118 identity condition. The same was true across subjects (**Fig. 3c**; Wilcoxon one-side signed-rank test,

119 statistics $= 28$, $p < 0.01$) revealing a systematic and consistent improvement of performance in the identity

120 condition.

7

121   The same group of subjects were also tested in the I(FP) task, and for the same three gains. As evident

122   from the behavior of the same typical subject, RMSE was smaller when the measurement and production

123   intervals were the same (**Fig. 3d and 3e**) compared to when the measurement period was longer (G =

124   0.667, t_{198} = 29.7316, p < 0.001, hedges' g = 4.1887), or shorter (G = 1.6, t_{198} = 25.6390, p < 0.001,

125   hedges' g = 3.6122). This effect was present across subjects (**Fig. 3f**; Wilcoxon one-side signed-rank test,

126   statistics = 28, p < 0.01) indicating that spatiotemporal identity helped subjects improve their estimate of

127   TTC. We also compared subjects' performance in the identity condition between the I(FP) and I(VM)

128   conditions (**Supplementary Fig. 3a**). RMSE was consistently and significantly smaller in the I(VM)

129   condition (Wilcoxon one-side signed-rank test, statistics = 3, p < 0.001). This ruled out the possibility that

130   subjects switched to a pure timing strategy in the identity context. These results suggest that subjects

131   exploited the temporal structure to improve their performance.

Experiment 3: Interception performance improves with temporal-identity context

132   Experiment 2 clearly demonstrated that interception was most accurate in the identity context when the

133   visible and occluded parts of the path were identical. This is consistent with our hypothesis that

134   performance benefited from the fact that the visible and occluded intervals had the same duration (i.e.,

135   temporal identity) allowing subjects to more accurately estimate TTC. However, it is also possible that

136   this improvement was because the visible and occluded parts had the same length (i.e., distance identity)

137   allowing subjects to better estimate distance. The latter hypothesis seems unlikely given that the occluded

138   distance was fixed throughout all experiments. Nonetheless, we conducted an additional experiment to

139   assess the relevance of temporal versus distance identity in interception performance.

140   Since distance and duration are related through speed, the only way to dissociate the two is to make the

141   speed of the bar differ between the visible and occluded parts of the path. Therefore, we designed a

142   variant of the interception task in which, unbeknownst to the subjects, the speed of the bar behind the

143   occluder was made 1.25 times faster than the speed in the visible portion (**Fig. 4a**). The non-identical

8

144   speed ratio enables us to create conditions in which the distance and temporal identity were dissociated.

145   In one condition, the visible and occluded distances were the same creating distance identity ($G_d = 1$)

146   without temporal identity ($G_t = 1/1.25$). In another condition, we matched the ratio of the distances to the

147   ratio of the speeds ($G_d = 1.25$) to create temporal identity ($G_t = 1$).

148   A new set of subjects was recruited for this experiment to ensure that any sensitivity to temporal context

149   was not because of participation in previous experiments. Since subjects were not aware of the speed

150   change behind the occluder, they could only adjust their performance based on feedback. We compared

151   subjects' performance between the $G_d = 1$ and $G_t = 1$ conditions. We reasoned that an observer that relies

152   on the distance identity should have higher performance (lower RMSE) in the $G_d = 1$ condition. In

153   contrast, an observer that relies on the temporal identity would have a lower RMSE in the $G_t = 1$ despite

154   the fact that the distances between the visible and occluded parts are not the same.

155   We found that RMSE was lower for the temporal identity compared to distance identity condition as

156   shown for a typical subject (**Fig. 4b and 4c**; $t_{198} = 25.6431$, $p < 0.001$, hedges' $g = 3.6127$) and across

157   subjects (Wilcoxon one-side signed-rank test, statistics = 34, $p < 0.05$). This finding further substantiates

158   our conclusion that subjects rely more on temporal context to estimate TTC.

Bayesian integration of speed and time explains interception performance

159   Experiments 1 to 3 established that subjects integrate speed and timing information to improve their

160   performance. Another salient feature of subjects' behavior across all conditions (**Fig. 2** and **Fig. 3** and

161   **Fig. 4**), was that *TTCp* was biased toward the mean of the prior. This was true for the external timing cue

162   tasks in Experiment 1, for the inherent timing tasks in Experiment 2, and in the control condition in

163   Experiment 3, regardless of the ratio of the occluded to visible lengths (different values of G). Together,

164   these observations suggest that subjects performance may be explained by a Bayesian model that

165   integrates the prior with both the speed and timing information (**Fig. 5a**).

166　　To test this hypothesis rigorously, we developed an ideal Bayesian observer for the task. We assumed that

167　　the ideal observer made two conditionally independent measurements while the moving bar was visible,

168　　one associated with the speed of the bar (v_m), and another associated with the duration of the visible

169　　interval (t_m). Following previous work, we assumed that these measurement were subject to scalar

170　　variability [26,33–36]. In particular, we assumed that the standard deviation of noise on speed scaled with the

171　　bar's speed (v) with constant of proportionality (w_{mV}) and standard deviation of noise on elapsed time

172　　scaled with visible duration (t_1) with constant of proportionality (w_{mT}). The ideal observer integrated the

173　　prior, $p(t_1)$, with the likelihood of the bar speed, $\lambda(v_m|v)$ and the likelihood of the visible duration, $\lambda(t_m|t_1)$,

174　　and computed TTCe from the mean of the posterior. Since this observer minimizes the least-squares error,

175　　we will refer to this as the Bayes least-squares (BLS) estimator. To compare the model to subjects'

176　　behavior, we augmented the ideal observer with a production stage by adding scalar noise with constant

177　　of proportionality (w_p) to *TTCe* to values of *TTCp* that incorporated motor variability.

178　　We first estimated w_{mV}, w_{mT} for each subject. In most Bayesian models, the model is evaluated by

179　　assessing the quality of model fits to the data. A more powerful approach is to fit the model to a training

180　　dataset and examine how well it explains a test dataset. An even more powerful approach is to fit the

181　　model to one set of conditions and ask whether it predicts data in another condition to which it was not

182　　fitted. We employed the last approach. For each subject, we estimated w_{mT} from the I(FP) task with G = 1

183　　(**Fig. 5b, left**), and w_{mV} from I(VM+FP) in G = 0.667 (**Fig. 5b, right**), and used those estimates to predict

184　　subjects' behavior in the I(VM+FP) in G = 1 (**Fig. 5c**).

185　　To estimate w_{mT}, we developed a Bayesian observer for the I(FP) task with G = 1. In this task, the sensory

186　　information provided was the interval between when the bar started to move and when it reached halfway

187　　along the path (over the fixation point), which we denote by t_1. We fitted subjects' behavior by a BLS

188　　estimator that only relied on the likelihood of t_1, $\lambda(t_m|t_1)$ and the prior distribution, $p(t_1)$. As shown for one

10

189   subject (**Fig. 5b, left**) and consistent with previous work in a similar task[21, 37–39], the model accurately

190   captured subjects' behavior.

191   Next, we estimated w_{mV} from fits of the Bayesian model to the I(VM+FP) task when G = 0.667. For this

192   fitting procedure, we used the corresponding w_{mT} from the I(FP) task with G = 0.667 (see **Methods**). As

193   shown for the same subject (**Fig. 5b, left**), the model successfully accounted for the behavior. Recall that

194   in the I(VM+FP) task, we had made the visible length long enough so that subjects' estimate of speed had

195   saturated and was thus no longer dependent on G (**Supplementary Fig. 2**). This allowed us to safely use

196   the fit to w_{mV} derived from the G = 0.667 condition to predict behavior in the G = 1 condition.

197   Finally, we use each subject's fits to w_{mV}, w_{mT} to predict the behavior in I(VM+FP) task when gain is one

198   (G = 1). The model was able to predict the observed *TTCp* values as shown for one example subject (**Fig.**

199   **5c**) and captured the data's summary statistics (BIAS and VAR) across subjects (**Fig. 5d**). This is

200   remarkable considering that both w_{mV} and w_{mT} were estimated from other tasks, and provides strong

201   support that subjects integrate prior information, speed information, and timing information to optimize

202   their estimate of TTC.

203   To further evaluate the success of the Bayesian model in explaining how subjects integrate speed and

204   timing information, we tested the model in Experiment 1 where the timing cue was provided explicitly by

205   a flash at the fixation point. To create a predictive model, we used the same procedure as we did to predict

206   behavior in Experiment 2. We estimated w_{mV} from data in I(VM) condition, and w_{mT} from data in I(FP)

207   condition (**Supplementary Fig. 3b**), and used those values to predict behavior in I(VM+FP) condition

208   (**Supplementary Fig. 1b**). Again, the model successfully captured the statistics of subjects' behavior

209   suggesting that the brain is optimized for integrating speed and timing information during interception

210   tasks regardless of how timing information is provided.

Discussion

211   Our work builds on a large body of work investigating the computational principles of object interception.

212   Early studies hypothesized that humans rely on variables derived from an object's visual angle and its rate

213   of expansion on the retina, of which the so-called tau is a classic example[1–3]. Later, this proposal was

214   deemed inadequate as it failed to capture many empirical observations[4, 9, 40–42]. Most current models are

215   based on the idea that interception relies on measurements of kinematic variables[6, 7, 11, 12, 43], such as

216   speed[6, 32, 44], distance and/or depth[45]. This idea has also been used in experiments similar to ours where the

217   object moves behind an occluder[6, 9, 46–48]. In those cases, it is assumed that humans estimate speed while

218   the object is visible and use that estimate to predict future position of the object behind the occluder. This

219   focus on kinematics is natural as it matches our intuition about the physics of how objects move.

220   However, the algorithms the brain uses for object interception need not match our physics intuition. Here,

221   we asked whether humans solely rely on kinematics (e.g., speed and distance), or do they additionally rely

222   on temporal cues and contexts.

223   Real world object interception involves a decision to initiate a movement followed by online adjustments

224   of the movement based on sensorimotor feedback. Although successful interception requires a tight

225   coordination between the initiation and the subsequent adjustments, the two processes typically involve

226   different computations[49]. The decision of when to initiate is, by and large, determined by a prediction of

227   how long it would take to reach the object – i.e., time-to-contact (TTC), whereas the subsequent

228   adjustment involves fine adjustments after the movement has been initiated. Here, we focused on the

229   former asking how the brain determines TTC. To do so, we designed a virtual interception task in which

230   subjects "intercepted" a moving bar by pressing a button when the bar reached a target position behind an

231   occluder. With this design, we effectively eliminated the need for post-initiation adjustments. Our main

232   objective was to investigate whether TTC was computed solely from estimates of bar kinematics (e.g.,

233   speed and distance), or whether subjects additionally relied on temporal cues and contexts.

12

234   We tested this question in two complementary sets of experiments. In the first set, we presented a brief

235   flash showing the position of the bar after its disappearance behind the occluder. This flash provided

236   explicit information about position and time of the bar but not its speed. Therefore, any improvement in

237   performance due to the flash must be taking advantage of timing mechanisms in the brain. Results

238   confirmed that subjects could intercept the bar without any speed information, and when the flash was

239   presented along with additional speed information, subjects were able to integrate the two to improve

240   their performance. This result complements a large body of evidence that humans are able to fuse sensory

241   information from multiple modalities while making perceptual inferences[29, 31, 50]. Note that the integration

242   of speed and time is distinct from the indirect role that time would play by improving one's estimate of

243   speed[51–53]. As we demonstrated in a supporting experiment (**Supplementary Fig. 2**), the improvement of

244   speed estimate with time saturates rapidly and cannot account for our finding. What our results reveal is

245   that humans actively use elapsed time as an independent cue and integrate it with other visual cues when

246   interacting with dynamic stimuli.

247   In the second set of experiments, we removed the explicit timing cue and instead asked whether subjects

248   would naturally exploit implicit timing cues present in the temporal structure of the environment. To

249   address this question, we designed an interception task in which we varied the interval the bar was visible.

250   Based on recent work[54], we reasoned that when the visible and occluded epochs have the same duration,

251   subjects would automatically make use of this temporal identity to improve their performance. Subjects'

252   performance was remarkably improved in the temporal identity context compared to when the durations

253   of the visible and occluded regions were not the same. Indeed, this experiment revealed a surprising

254   aspect of human behavior: performance in the identity context was even better than when the occluded

255   length was the same and the visible length was made longer. In other words, prolonging the visible

256   portion was harmful to performance when it broke the temporal structure conferred by the identity

257   context. This result powerfully demonstrated that the key factor driving the performance improvement

258   was the presence of the identity context. This conclusion was reinforced by control experiments showing

259   that the result was due to temporal – not distance – identity. Finally, we found that subjects' ability to

260   integrate speed and timing information reached performance levels similar to an ideal Bayesian observer,

261   suggesting that human brain is inherently optimized to combine speed and time information for object

262   interception.

263   These experiments lead to a simple and novel conclusion that humans actively engage timing mechanisms

264   during interception. To put this finding in context, it is important to distinguish between the role of time

265   during the visible and occluded regions of the path. When an object moves behind an occluder, subjects

266   could no longer measure the object's speed and thus have no choice but to rely on their sense of time.

267   This idea was formalized by Tresillian and others in relation to human's ability to extrapolate an object's

268   location behind an occluder[55, 56]. This is fundamentally different from what we propose; our findings

269   indicate that humans actively integrate information about temporal contexts and events even when the

270   object is visible. In other words, timing seem to be an integral component of how we interact with

271   dynamics stimuli, both to better estimate where they are (when they are visible), and to infer where they

272   might be (when they are occluded).

273   Our work does not address any potential role that timing information might play for the subsequent

274   sensorimotor adjustments after movement initiation. It is possible that knowledge about temporal cues

275   and contexts only inform movement initiation. This would indicate that temporal processing is only

276   engaged during the cognitive and/or motor planning stage of object interception. This is consistent with

277   numerous imaging and electrophysiological studies finding an important role for premotor and

278   supplementary motor areas in timing[57–61]. Alternatively, knowledge about movement durations may also

279   be used during movements although some studies have suggested that humans do not use timing

280   information when they have access to movement related state-dependent information[62, 63].

281   It is worthwhile considering why the role of time was not noted in prior research on object interception.

282   We think that answer has to do with the simplicity of behavioral task used in laboratory settings (but see

283   some studies using more naturalistic paradigms or done with virtual reality[47, 64]). Most experiments have

284   not included rich spatiotemporal event and/or contexts, where temporal cues become relevant. However,

285   real world examples of object interception take place in the presence of temporal statistics, spatial

286   landmarks, and temporal events such as collisions and/or reflections, all of which make knowledge about

287   time highly informative. A notable observation in our experiment was that subjects' estimate of TTC was

288   more accurate in the identity temporal context, possibly due to lower sensorimotor noise[54]. This improved

289   sensitivity may be due to the fact that temporal identity creates a rhythmic structure between the relevant

290   time intervals. If so, we would expect stronger effects when temporal events create sounds as auditory

291   rhythms and/or integer ratios are constrained by strong internal priors[65]. For example, intercepting a

292   bouncing ball may greatly benefit from the bounce sound, especially when visual information is uncertain

293   (e.g., a dribbling a basketball without looking at the ball). These considerations highlight the need for

294   future research to move beyond simple behavioral tasks and examine object interception in more

295   naturalistic settings where the underlying dynamics are governed by richer spatiotemporal contexts. We

296   speculate that doing so will further substantiate the importance of temporal events and contexts in

297   processing dynamic stimuli.

15

Online Methods

Subjects

298   All subjects provided informed consent for experimental procedures which were approved by the

299   Committee On the Use of Humans as Experimental Subjects at the Massachusetts Institute of

300   Technology. Seven adult subjects participated in Experiment 1. A different group of seven adult subjects

301   participated in Experiment 2. Another group of eight adult subjects participated in Experiment 3. All

302   subjects had normal or corrected-to-normal vision.

Procedures

303   Subjects sat in a dark, quiet room at a distance of approximately 50 cm from a display monitor with a

304   refresh rate of 60 Hz and a resolution of 1920 by 1200 on an Apple Macintosh platform. Experiments

305   were controlled by an open-source software (MWorks; http://mworks-project.org/). All stimuli were

306   presented on a black background. Although eye movements were not monitored, all trials began with

307   central fixation spot that subjects were asked to hold their gaze on the fixation point throughout every

308   trial. Responses were made on a standard Apple Keyboard connected to the experimental machine.

309   We used three experiments to examine how people infer time-to-contact (TTC). Each experiment

310   consisted of conditions whose order was randomized across subjects. Each condition was tested twice in

311   two different days: the first session was used for training, and the second was used for the main test

312   session but the first 25 trials were considered as warm-up and were excluded from the main analysis.

313   Subjects were asked to press a key when the bar reached the target position. Feedback was provided to

314   indicate the actual bar position along the path when the key was pressed. The target position and the

315   stimulus feedback were shown in green when the produced time-to-contact (*TTCp*) was within an

316   experimentally defined window around the actual time-to-contact (*TTCa*), and red otherwise. To account

16

317 for scalar variability, the window width was scaled with the actual time-to-contact (*TTCa*) with a constant

318 of proportionality, *k*. The value of *k* was controlled by an adaptive one-up one-down procedure during

319 training condition, and eventually reached a stable value, k_0. We then set k_0 as feedback accuracy window

320 in the main test session.

Experiment 1

321 The objective of this experiment was to test whether subjects could improve their estimate of TTC by

322 integrating motion and timing cues. The trials were structured as follows: subject pressed a key to initiate

323 a trial. After a variable delay drawn randomly from a truncated exponential distribution (0.3-0.6 sec), a

324 bar started moving horizontally from a starting point along a 16-degrees long linear path toward a target

325 position at the end of the path (**Fig. 1a**). In each trial, the speed (v) of the bar was drawn from a discrete

326 uniform distribution (**Fig. 1b**).

327 The experiment consisted of three different conditions in terms of the information subjects were provided

328 with: one with motion cue, one with timing cue, and one with both (**Fig. 2a**). In the first condition, the

329 interception path consisted of two sections: a section where the stimulus movement was visible and a

330 section where it was occluded. The target was placed at the end of the occluded section. We denote this

331 condition by I(VM) as a shorthand for Interception in the presence of Visible Motion. In the second

332 condition, the motion was invisible throughout the path but the position of the stimulus was flashed at the

333 beginning of the path and when it reached the central fixation point in the middle of the path. We denote

334 this condition by I(FP) as a shorthand for Interception in the presence of Flashed Position. In the third

335 condition, the stimulus movement was visible early on, and additionally its position was flashed when it

336 reached the central fixation spot. Accordingly, we denote this condition by I(VM+FP). In tasks in which

337 the position was flashed, the flashes lasted 100 msec. The distribution of sample interval (t_1) between the

338 start of the path and the time when the bar reached the fixation point was the same across the conditions.

17

339   We also tested subjects in a cue conflict version of the I(VM+FP) in which the flash at the central fixation

340   was jittered by –100, 0, or 100 msec relative to when the actual stimulus reached the central

341   fixation (**Supplementary Fig. 1a**). These three jitter values were randomized and presented with equal

342   probability.

Experiment 2

343   The objective of this experiment was to test whether subjects could take advantage of temporal structure

344   in the absence of an explicit temporal cue to improve their performance. The task was similar to the

345   I(VM) condition in Experiment 1. A bar began to move from a starting point along a path. The bar was

346   initially visible and then disappeared behind an occluder. Subjects pressed a key when the bar reached the

347   target position at the end of the occluder. We tested subjects in three conditions (**Fig. 3a**). In all

348   conditions, the distance between the fixation spot and target (d_2) was set at 8 degrees while the visible

349   length between the starting and the fixation spots (d_1) was varied between 12, 8, and 5 degrees. We

350   expressed these conditions in terms of the ratio of d_2 over d_1, which we define as a gain factor (G). The

351   corresponding G for the three conditions were 0.667, 1, or 1.6. We recruited a new set of subjects for this

352   experiment to make sure that participants were not made sensitive to timing cues due to prior experience

353   with the I(FP) and/or I(VM+FP) tasks.

354   To evaluate the relative importance of speed and timing information, we also tested the newly recruited

355   subjects with the same gain factors but in the I(FP) condition (**Fig. 3d**). However, all I(FP) conditions

356   were tested after the subjects had completed the I(VM) conditions to avoid inadvertently sensitizing

357   subjects to timing cues. Overall, Experiment 2 consisted of 6 conditions in total.

Experiment 3

358   The objective of this experiment was to test whether the improved performance in Experiment 2 in

359   relation to the identity context was related to the distance identity or temporal identity. To facilitate the

360   description of these conditions, let us define the sample interval (t_1) as the interval associated with the

361   visible portion of the path, between the starting point and when the bar reached the central fixation (d_1).

362   Similarly, we define the target interval (t_2) as the interval associated with the occluded part of the path

363   (d_2). Since the experiment involved changing the relative distances and/or durations, we additionally

364   define two a distance ratio that corresponds to the ration of the occluded length to the visible length ($G_d =$

365   d_2/d_1), and a duration ratio ($G_t = t_2/t_1$) for the corresponding durations.

366   The experiment consisted of two variants of the I(VM) condition. In the first condition, we set d_1 to 8

367   degrees and d_2 to 10 degrees (G_d=1.25), and in the second condition both d_1 and d_2 were 10 degrees (G_d

368   =1). In the training sessions, similar to experiment 1 and 2, the stimulus speed was constant throughout

369   the interception path. In the test sessions, unbeknownst to the subjects, immediately after the stimulus

370   entered the occluded segment, its speed was multiplied by 1.25. This manipulation changed G_t to 1 and

371   1/1.25 in the first and second conditions respectively. In other words, in the first condition, G_d=1.25 and

372   G_t=1, whereas in the second condition G_d=1 and G_t=1/1.25. This manipulation allowed us to tease apart

373   the effect of distance and temporal identity contexts (**Fig. 4a**).

Analysis

374   We defined the actual time-to-contact (*TTCa*) as the interval between when the bar passed the central

375   fixation to when it reached the target position. The produced time-to-contact (*TTCp*) was defined as the

376   time from when the bar passed the central fixation to when the subject pressed a key. Subjects that were

377   not sensitive to the range of sample intervals during the training session or had unstable performance were

378   excluded from the study. We considered a subject insensitive if the corresponding *TTCp* distribution for

379   the longest *TTCa* was not significantly different from *TTCp* distribution for the shortest *TTCa* (paired t-

380   test at p=0.05 level). Performance was considered unstable if the first and second order statistics of *TTCp*

381   were different between the first and second halves of the session (paired t-test at p=0.05 level).

19

Summary statistics

382   Following Jazayeri and Shadlen (2010), we characterized each subject's performance by computing the

383   following summary statistics for *TTCp*:

$$\text{BIAS}^2 = \frac{1}{5} \times \sum_{i=1}^{5} \text{BIAS}_i^2 = \frac{1}{5} \times \sum_{i=1}^{5} (\overline{TTCp}_i - TTCa_i)^2$$

$$\text{VAR} = \frac{1}{5} \times \sum_{i=1}^{5} \text{VAR}_i = \frac{1}{5} \times \sum_{i=1}^{5} \left(\frac{1}{N_i - 1} \sum_{k=1}^{N_i} (TTCp_{ik} - \overline{TTCp}_i)^2 \right)$$

$$\text{RMSE} = \sqrt{\text{BIAS}^2 + \text{VAR}}$$

384   BIAS and VAR represent the average deviations and average variance over the five intervals included in

385   the prior distribution. BIAS_i and VAR_i represent the mean deviation and variance of produced times

386   (*TTCp*) for the *i*-th actual interval (*TTCa*) with N_i trials. It follows naturally that the overall root mean

387   squared error (RMSE) is equal to the square root of the sum of BIAS^2 and VAR. To estimate the mean

388   and variance of summary statistics for individual subject in each condition, we resampled data with

389   replacement and repeated this resampling 100 times

Effect size

390   It is known that a relatively large sample size could lead to a smaller p value. Since there were more than

391   100 trials in each session for each subject, we also measured the strength of difference between conditions

392   for each subject. We used Hedges' g[66], which is a measure to correct the bias in Cohen's *d* to quantify the

393   distance between two distribution means. $g = 0.2$ means small effect size, 0.5 means medium effect size,

394   and 0.8 means large effect size.

$$g = J(n_1 + n_2 - 2) \cdot \frac{(\bar{x}_1 - \bar{x}_2)}{s*}$$

395   where

$$s* = \frac{\sqrt{(n_1 - 1) \cdot s_1{}^2 + (n_2 - 1) \cdot s_2{}^2}}{n_1 + n_2 - 2} \qquad J(a) = \frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\frac{a}{2}} \cdot \Gamma\left(\frac{a-1}{2}\right)}$$

20

The Bayesian observer model

396 We developed a Bayesian observer model (**Fig. 5a**) based on previous work on interval reproduction[24].

397 We modeled the prior distribution over sample intervals (t_1) based on the ratio of the visible distance (d_1)

398 to the bar's speed (v). To simplify derivations, we modeled the discrete prior distributions used in the

399 experiment as a continuous uniform distribution ranging from the shortest to longest sample interval tests.

400 The shortest and longest intervals were computed in terms of the smallest and largest speeds ($v^{\min}$ and

401 $v^{\max}$).

$$p(t_1) = \begin{cases} \dfrac{v^{\min} \cdot v^{\max}}{d_1 \cdot (v^{\max} - v^{\min})} & \dfrac{d_1}{v^{\max}} \leq t_1 \leq \dfrac{d_1}{v^{\min}} \\ 0 & \text{otherwise} \end{cases}$$

402 We assumed that subjects made two conditionally independent measurements when the bar was visible,

403 one associated with the speed of the bar, and another associated with duration of the visible period.

404 Following previous work on sensory measurements of time and speed[26, 33–36], we assumed that both

405 measurements were perturbed by scalar Gaussian noise. Specifically, we assumed that the standard

406 deviation of measured speed (v_m) scales with speed (v) with constant of proportionality w_{mV}, and that the

407 standard deviation of measured elapsed time (t_m) scales with elapsed time (t_1) with constant of

408 proportionality w_{mT}. The variables w_{mV} and w_{mT} represent the Weber fraction for measurement of speed

409 and time, respectively. From the perspective of the observer who makes a measurement t_m and v_m, but

410 does not know t_1 and v, the problem can be written in terms of the corresponding likelihood functions

411 $\lambda(t_m | t_1)$ and $\lambda(v_m | v)$:

$$\lambda(t_m | t_1) = \frac{1}{\sqrt{2\pi \cdot (w_{mT} \cdot t_1)^2}} \times e^{\frac{-(t_m - t_1)^2}{2 \cdot (w_{mT} \cdot t_1)^2}}$$

$$\lambda(v_m | v) = \frac{1}{\sqrt{2\pi \cdot (w_{mV} \cdot v)^2}} \times e^{\frac{-(v_m - v)^2}{2 \cdot (w_{mV} \cdot v)^2}}$$

412   To be able to combine the two likelihoods, we rewrote the likelihood associated with speed in terms of

413   distance and sample interval, as follows:

$$\lambda(v_m | t_1) = \frac{1}{\sqrt{2\pi \cdot \left(w_{mV} \cdot \frac{d_1}{t_1}\right)^2}} \times e^{\frac{-\left(v_m - \frac{d_1}{t_1}\right)^2}{2 \cdot \left(w_{mV} \cdot \frac{d_1}{t_1}\right)^2}}$$

414   Assuming that the two measurements were conditionally independent, the posterior would be

$$\pi(t_1 | t_m, v_m) = \frac{p(t_1) \cdot p(t_m | t_1) \cdot p(v_m | t_1)}{\int p(t_1) \cdot p(t_m | t_1) \cdot p(v_m | t_1) \ dt_1}$$

415   Following previous work[24], we assumed that subjects' minimized expected loss using a quadratic loss

416   function, and modeled the inferred duration of the visible period based on the Bayes least-squares (BLS)

417   estimator (i.e., mean of the posterior). We assumed that the estimate was multiplied with a lossless gain

418   (G) to get time-to-contact compute (*TTCe*).

$$
\begin{aligned}
TTC_e &= G \cdot f_{\mathrm{BLS}}(t_m, v_m) \\
&= G \cdot \underset{t_e}{\mathrm{argmax}} \left[\int (t_e - t_1)^2 \cdot \pi(t_1 | t_m, v_m) \ dt_1 \right] \\
&= G \cdot \int t_1 \cdot \pi(t_1 | t_m, v_m) \ dt_1
\end{aligned}
$$

419   For a uniform prior for sample interval (t_1), the time-to-contact estimate (*TTCe*) would be

$$TTC_e = G \cdot \frac{\int_{t_{1,min}}^{t_{1,max}} t_1 \cdot p(t_m | t_1) \cdot p(v_m | t_1) \ dt_1}{\int_{t_{1,min}}^{t_{1,max}} p(t_m | t_1) \cdot p(v_m | t_1) \ dt_1}$$

420   The model was augmented by post-estimation noise to account for motor variability in the produced time-

421   to-contact (*TTCp*). Following previous work[24, 33, 34, 67], we assumed that the standard deviation of motor

422   noise was proportional to *TTCe*, with constant of proportionality of w_p (Weber fraction for production).

423   We included an offset term (b_0) in the fitting procedure to account for idiosyncratic stimulus- and prior-

424   independent biases observed in responses.

22

$$p(TTC_p|TTC_e) = \frac{1}{\sqrt{2\pi \cdot (w_p \cdot TTC_e)^2}} \times e^{\frac{-(TTC_p - (TTC_e + b_0))^2}{2 \cdot (w_p \cdot (TTC_e + b_0))^2}}$$

425   Using chain rule and marginalization of hidden variables, we wrote the the conditional probability of

426   produced time-to-contact (*TTCp*) for a each actual time-to-contact (*TTCa*) as follows:

$$p(TTC_p|TTC_a; w_{mT}, w_{mV}, w_p, b_0)$$
$$= \iint p(TTCp|G_d \cdot f_{\mathrm{BLS}}(t_m, v_m); w_p, b_0) \cdot p(t_m|(TTC_a/G_d); w_{mT}) \cdot p(v_m|(TTC_a/G_d); w_{mV}) \ dt_m \ dv_m$$

Fitting procedure

427   For fitting procedure, we assumed that *TTCp* values associated with any *TTCa* were independent across

428   trials and thus expressed the joint conditional probability of individual *TTCp* values across all the N trials

429   by the product of their individual conditional probabilities.

$$\log p(TTC_p^1, TTC_p^2, ..., TTC_p^N|TTC_a; w_{mV}, w_{mT}, w_p, b_0) = \sum_{i=1}^{N} \log p(TTC_p^i|TTC_a; w_{mV}, w_{mT}, w_p, b_0)$$

430   We used fminsearch algorithm to find the model parameters that maximized the likelihood of model

431   parameters across all *TTCa* and *TTCp* values measured psychophysically. Integrals were approximated

432   numerically using the global adaptive quadrature[68]. We repeated the search with different initial values 10

433   times, and verified that the likelihood functions were stable with respect to initial values.

Predicting behavior in temporal identity context

434   Instead of fitting the Bayesian model to each dataset, we asked whether we could fit the model to some

435   conditions and then use parameters of the fitted model to *predict* behavior in other conditions. We aimed

436   to predict behavior in the most important condition where subjects integrated speed with the identity

437   temporal context; i.e., I(VM) with G = 1. We assumed that the noise associated with the measurement of

438   t_1 is the same in the I(VM) and I(FP) tasks and therefore, used the Bayesian model to the I(FP) task for G

439   = 1 to estimate w_{mT} (**Fig. 5b, left**). We further assumed that the measurement of speed in I(VM) task

23

440   would be the same across two different gains (G = 1 and G = 0.667), given that the accuracy of speed

441   measurement saturated rapidly (**Supplementary Fig. 2**). We first found w_{mT} for G = 0.667 from I(FP)

442   task, and then used this value to fit a Bayesian model to I(VM) task with G = 0.667 to estimate w_{mV} (**Fig.**

443   **5b, right**). Finally, we used the w_{mT} inferred from I(FP) with G = 1 and w_{mV} inferred from I(VM) with G

444   = 0.667 to predict behavior in the I(VM) task with G = 1 in (**Fig. 5c**).

References

1. Bootsma, R. J. & Oudejans, R. R. Visual information about time-to-collision between two objects. *J. Exp. Psychol. Hum. Percept. Perform.* **19,** 1041–1052 (1993).

2. Lee, D. N., Young, D. S., Reddish, P. E., Lough, S. & Clayton, T. M. Visual timing in hitting an accelerating ball. *Q. J. Exp. Psychol. A* **35,** 333–346 (1983).

3. Savelsbergh, G. J., Whiting, H. T. & Bootsma, R. J. Grasping tau. *J. Exp. Psychol. Hum. Percept. Perform.* **17,** 315–322 (1991).

4. Smeets, J. B., Brenner, E., Trébuchet, S. & Mestre, D. R. Is judging time-to-contact based on 'tau'? *Perception* **25,** 583–590 (1996).

5. Brenner, E. & Smeets, J. Intercepting moving objects: do eye-movements matter. *Space and time in perception and action* (2010).

6. Kwon, O.-S. & Knill, D. C. The brain uses adaptive internal models of scene statistics for sensorimotor estimation and planning. *Proceedings of the National Academy of Sciences* **110,** E1064–E1073 (2013).

7. Mrotek, L. A., Flanders, M. & Soechting, J. F. Interception of targets using brief directional cues. *Exp. Brain Res.* **156,** 94–103 (2004).

8. Mrotek, L. A. & Soechting, J. F. Target interception: hand-eye coordination and strategies. *J. Neurosci.* **27,** 7297–7309 (2007).

9. Zago, M., McIntyre, J., Senot, P. & Lacquaniti, F. Visuo-motor coordination and internal models for object interception. *Exp. Brain Res.* **192,** 571–604 (2009).

10. Bosco, G., Delle Monache, S. & Lacquaniti, F. Catching what we can't see: manual interception of occluded fly-ball trajectories. *PLoS One* **7,** e49381 (2012).

11. Brouwer, A.-M., Brenner, E. & Smeets, J. B. J. Hitting moving objects: is target speed used in guiding the hand? *Exp. Brain Res.* **143,** 198–211 (2002).

12. Soechting, J. F., Juveli, J. Z. & Rao, H. M. Models for the extrapolation of target motion for manual interception. *J. Neurophysiol.* **102,** 1491–1502 (2009).

13. McIntyre, J., Zago, M., Berthoz, A. & Lacquaniti, F. Does the brain model Newton's laws? *Nat. Neurosci.* **4,** 693–694 (2001).

14. Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. Noise in the nervous system. *Nat. Rev. Neurosci.* **9,** 292–303 (2008).

15. Ahrens, M. B. & Sahani, M. Observers exploit stochastic models of sensory change to help judge the passage of time. *Curr. Biol.* **21,** 200–206 (2011).

16. Brenner, E. & Smeets, J. B. J. Sources of variability in interceptive movements. *Exp. Brain Res.* **195,** 117–133 (2009).

17. Pellizzer, G. & Hedges, J. H. Motor planning: effect of directional uncertainty with continuous spatial cues. *Exp. Brain Res.* **154,** 121–126 (2003).

18. Scheidt, R. A., Dingwell, J. B. & Mussa-Ivaldi, F. A. Learning to move amid uncertainty. *J. Neurophysiol.* **86,** 971–985 (2001).

19. Ota, K., Shinya, M. & Kudo, K. Motor planning under temporal uncertainty is suboptimal when the gain function is asymmetric. *Front. Comput. Neurosci.* **9,** 88 (2015).

20. van Beers, R. J., Baraduc, P. & Wolpert, D. M. Role of uncertainty in sensorimotor control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357,** 1137–1145 (2002).

21. Miyazaki, M., Nozaki, D. & Nakajima, Y. Testing Bayesian models of human coincidence timing. *J. Neurophysiol.* **94,** 395–399 (2005).

22. Aschersleben, G. & Prinz, W. Synchronizing actions with events: the role of sensory information. *Percept. Psychophys.* **57,** 305–317 (1995).

23. Chen, Y., Ding, M. & Kelso, J. Long memory processes (1/f α type) in human coordination. *Phys. Rev. Lett.* (1997).

24. Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13,** 1020–1026 (2010).

25. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27,** 712–719 (2004).

26. Stocker, A. A. & Simoncelli, E. P. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* **9,** 578–585 (2006).

27. Körding, K. P., Ku, S.-P. & Wolpert, D. M. Bayesian integration in force estimation. *J. Neurophysiol.* **92,** 3161–3165 (2004).

28. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427,** 244–247 (2004).

29. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415,** 429–433 (2002).

30. Tassinari, H., Hudson, T. E. & Landy, M. S. Combining Priors and Noisy Visual Cues in a Rapid Pointing Task. *J. Neurosci.* **26,** 10154–10163 (2006).

31. Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. Combining sensory information: mandatory fusion within, but not between, senses. *Science* **298,** 1627–1630 (2002).

32. Gray, R. & Regan, D. Accuracy of estimating time to collision using binocular and monocular information. *Vision Res.* **38,** 499–512 (1998).

33. Gibbon, J., Church, R. M. & Meck, W. H. Scalar Timing in Memory. *Ann. N. Y. Acad. Sci.* **423,** 52–77 (1984).

34. Rakitin, B. C. *et al.* Scalar expectancy theory and peak-interval timing in humans. *J. Exp. Psychol. Anim. Behav. Process.* **24,** 15–33 (1998).

35. Welch, L. The perception of moving plaids reveals two motion-processing stages. *Nature* **337,** 734–736 (1989).

36. McKee, S. P., Silverman, G. H. & Nakayama, K. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Res.* **26,** 609–619 (1986).

37. Jazayeri, M. & Shadlen, M. N. A Neural Mechanism for Sensing and Reproducing a Time Interval. *Curr. Biol.* **25,** 2599–2609 (2015).

38. Acerbi, L., Wolpert, D. M. & Vijayakumar, S. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput. Biol.* **8,** e1002771 (2012).

39. Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M. & Burr, D. C. Optimal encoding of interval timing in expert percussionists. *J. Neurosci.* **32,** 1056–1060 (2012).

40. Tresilian, J. R. Visually timed action: time-out for 'tau'? *Trends Cogn. Sci.* **3,** 301–310 (1999).

41. Hecht, H. & Savelsbergh, G. J. P. in *Advances in Psychology* (ed. Heiko Hecht and Geert J.P. Savelsburgh) **Volume 135,** 1–11 (North-Holland, 2004).

42. López-Moliner, J., Field, D. T. & Wann, J. P. Interceptive timing: prior knowledge matters. *J. Vis.* **7,** 11.1–8 (2007).

43. Brenner, E. & Smeets, J. B. J. How people achieve their amazing temporal precision in interception. *J. Vis.* **15,** (2015).

44. Lacquaniti, F. & Maioli, C. The role of preparation in tuning anticipatory and reflex responses during catching. *J. Neurosci.* **9,** 134–148 (1989).

45. Hecht, H., Kaiser, M. K. & Banks, M. S. Gravitational acceleration as a cue for absolute size and distance? *Percept. Psychophys.* **58,** 1066–1075 (1996).

46. Lee, D. D., Ortega, P. A. & Stocker, A. A. Dynamic belief state representations. *Curr. Opin. Neurobiol.* **25,** 221–227 (2014).

47. Fooken, J., Yeo, S.-H., Pai, D. K. & Spering, M. Eye movement accuracy determines natural interception strategies. *J. Vis.* **16,** 1 (2016).

48. Dubrowski, A., Lam, J. & Carnahan, H. Target velocity effects on manual interception kinematics. *Acta Psychol.* **104,** 103–118 (2000).

49. Resulaj, A., Kiani, R., Wolpert, D. M. & Shadlen, M. N. Changes of mind in decision-making. *Nature* **461,** 263–266 (2009).

50. Jacobs, R. A. & Fine, I. Experience-dependent integration of texture and motion cues to depth. *Vision Res.* **39,** 4062–4075 (1999).

51. Cavallo, V. & Laurent, M. Visual information and skill level in time-to-collision estimation. *Perception* **17,** 623–632 (1988).

52. Mason, A. H. & Carnahan, H. Target viewing time and velocity effects on prehension. *Exp. Brain Res.* **127,** 83–94 (1999).

53. Krukowski, A. E., Pirog, K. A., Beutter, B. R., Brooks, K. R. & Stone, L. S. Human discrimination of visual direction of motion with and without smooth pursuit eye movements. *J. Vis.* **3,** 831–840 (2003).

54. Remington, E. & Jazayeri, M. Late Bayesian inference in sensorimotor behavior. *bioRxiv* 130062 (2017). doi:10.1101/130062

55. Tresilian, J. R. Perceptual and cognitive processes in time-to-contact estimation: analysis of prediction-motion and relative judgment tasks. *Percept. Psychophys.* **57,** 231–245 (1995).

56. DeLucia, P. R. & Liddell, G. W. Cognitive motion extrapolation and cognitive clocking in prediction motion task. *J. Exp. Psychol. Hum. Percept. Perform.* **24,** 901–914 (1998).

57. Merchant, H., Pérez, O., Zarco, W. & Gámez, J. Interval tuning in the primate medial premotor cortex as a general timing mechanism. *J. Neurosci.* **33,** 9082–9096 (2013).

58. Harrington, D. L., Zimbelman, J. L., Hinton, S. C. & Rao, S. M. Neural modulation of temporal encoding, maintenance, and decision processes. *Cereb. Cortex* **20,** 1274–1285 (2010).

59. Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y. & Tanji, J. Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nat. Neurosci.* **12,** 502–507 (2009).

60. Kunimatsu, J. & Tanaka, M. Alteration of the timing of self-initiated but not reactive saccades by electrical stimulation in the supplementary eye field. *Eur. J. Neurosci.* **36,** 3258–3268 (2012).

61. Cui, X., Stetson, C., Montague, P. R. & Eagleman, D. M. Ready...go: Amplitude of the FMRI signal encodes expectation of cue arrival time. *PLoS Biol.* **7,** e1000167 (2009).

62. Diedrichsen, J., Criscimagna-Hemminger, S. E. & Shadmehr, R. Dissociating timing and coordination as functions of the cerebellum. *J. Neurosci.* **27,** 6291–6301 (2007).

63. Conditt, M. A. & Mussa-Ivaldi, F. A. Central representation of time during motor learning. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11625–11630 (1999).

64. Diaz, G., Cooper, J., Rothkopf, C. & Hayhoe, M. Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *J. Vis.* **13,** (2013).

65. Jacoby, N. & McDermott, J. H. Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Curr. Biol.* **27,** 359–370 (2017).

66. Hedges, L. V. *Statistical Methods for Meta-Analysis*. (Academic Press, 1985).

67. Gallistel, C. R. & Gibbon, J. Time, rate, and conditioning. *Psychol. Rev.* **107,** 289–344 (2000).

68. Shampine, L. F. Vectorized adaptive quadrature in MATLAB. *J. Comput. Appl. Math.* **211,** 131–140 (2008).

Box

a

b
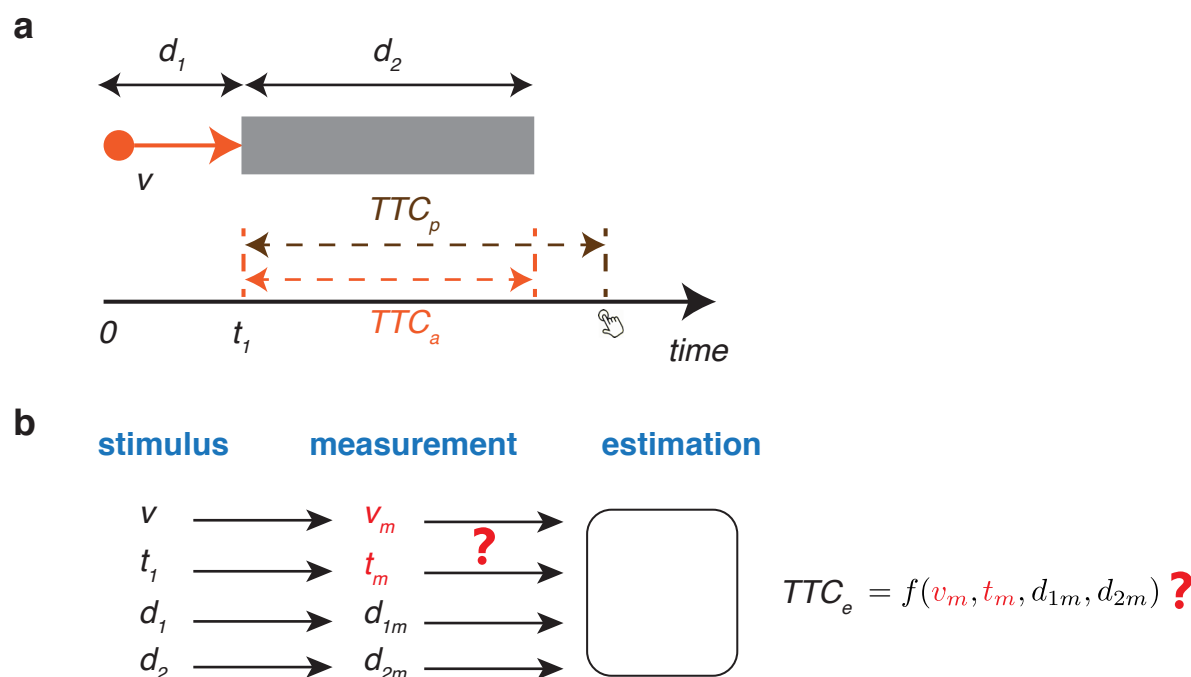
Box 1. Schematic paradigm and hypothesis for object interception.

(**a**) The overall logic of the experimental design. Subjects are asked to press a key when an bar moving with speed, v, would arrive at a target. The movement path is divided into a first section where the bar is visible, and a second section where the bar is invisible (i.e., virtually occluded). The subscript 1 and 2 are used to denote the distance (d), speed (v) and duration (t) of the two sections, respectively. Subjects' behavior is evaluated by comparing the actual time-to-contact ($TTCa$) to the produced time-to-contact ($TTCp$), both of which are measured with respect to the moment the bar goes behind the occluder. (**b**) An estimate of TTC (denoted $TTCe$) can be derived by applying an appropriate transformation (noted as function, f) to measured stimulus parameters (v_m, t_m, d_{1m}, d_{2m}). The key question we focus on is whether subjects rely solely on the speed (v_m), or they additionally incorporate temporal cues (t_m).

Figures
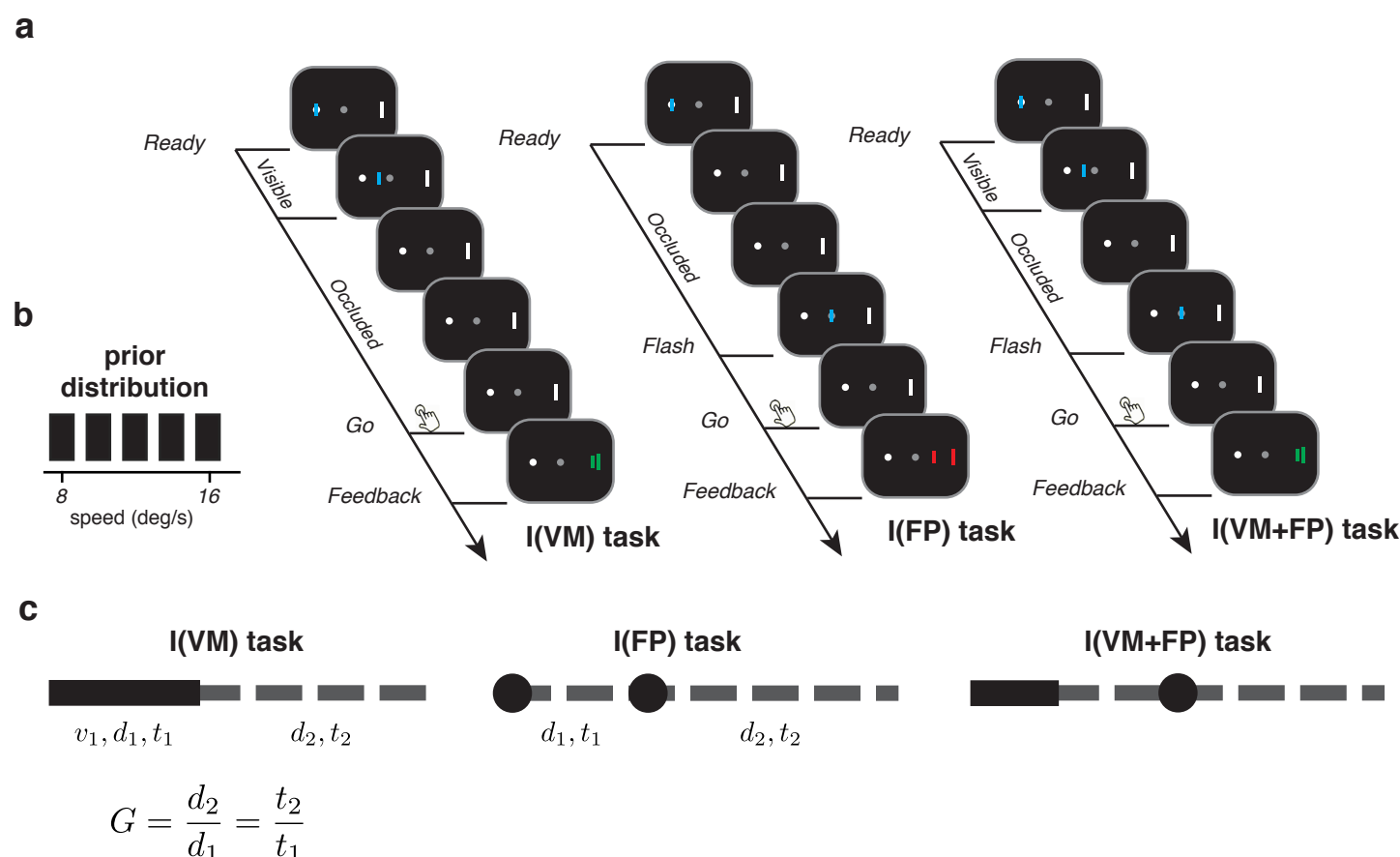
Figure 1. Experimental design and behavioral tasks.

(**a**) Behavioral tasks. A bar moved from an initial point to the left of the fixation point to a target point to the right of the fixation point. The initial and target points were present throughout the trial. Subjects had to press a key when they judged the moving bar to have arrived at the target. Using this basic design, we tested subjects behavior in three conditions. In the I(VM) task (Interception with Visual Motion), the stimulus movement was initially visible and then invisible as if behind an (imaginary) occluder. In the I(FP) task (Interception with Flashed Position), the motion was not displayed and the stimulus was only flashed at the starting point and when it reached the fixation point. In the I(VM+FP) task, both the initial movement and the intermediate flash at the fixation point were displayed. In all trials, we provided feedback by presenting the position of the stimulus at the time of keypress. To reinforce accuracy, the target position and the stimulus feedback were shown in green when $TTCp$ was within an experimentally defined window around $TTCa$, and red otherwise (see Methods). (**b**) Prior distribution of the stimulus speed. Speed was sampled from a discrete uniform distribution from 8 to 16 degree per second, and varied across trials. (**c**) Simplified symbolic representation of I(VM), I(FP), and I(VM+FP) conditions. The circles correspond to times when the stimulus was flashed, the solid lines to when the motion was displayed, and the dashed lines to when the stimulus was occluded. We used this symbolic representation as key for other figures.
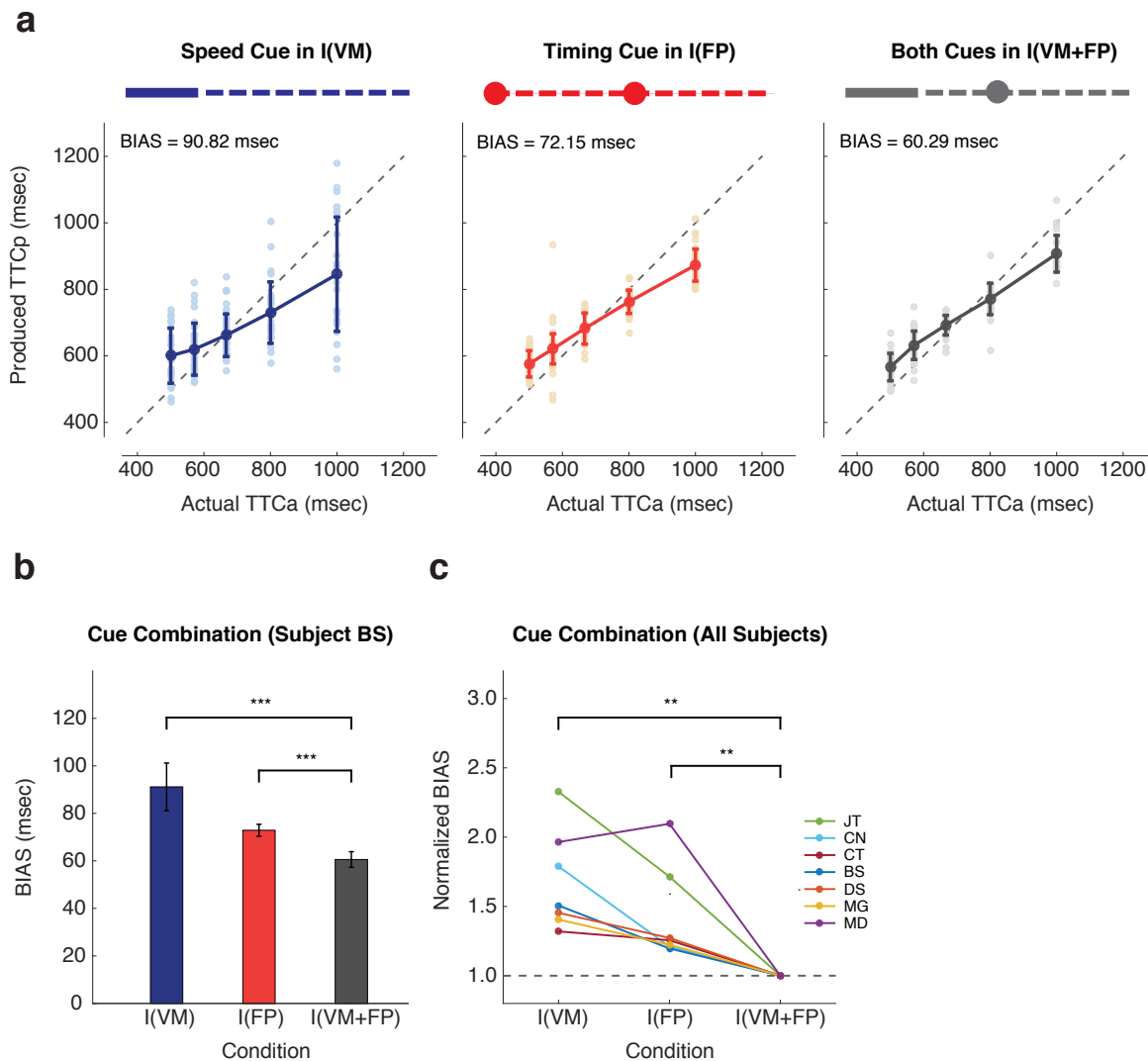
Figure 2. Interception using speed and explicit timing cues (Experiment 1).

(**a**) Behavior of a typical subject for different conditions in Experiment 1. The left panel corresponds to the I(VM) task were the speed of the stimulus was evident from the initial visible section of the movement. The middle panel corresponds to the I(FP) task in which the stimulus was flashed at the starting point and the central fixation point. The right panel corresponds to the I(VM+FP) task where both the initial speed and the intermediate flash were presented. Performance was quantified by comparing subject's produced time-to-contact (*TTCp*) to the actual time-to-contact (*TTCa*). *TTCa* was defined as the time between when the bar reached the central fixation to when it reached the target. *TTCp* was defined as the time between when the bar reached the central fixation to when the button was pressed. Light dots and dark circles show *TTCp* in each trial and the corresponding averages for each *TTCa*. BIAS in each plot was quantified as the average error over the five distinct *TTCa* of the prior distribution, i.e., the root mean square of differences between five solid dark circle and the corresponding diagonal dash line on the plot. (**b**) BIAS comparison across conditions for a typical subject. We estimated the standard error through resampling data with 100 repetitions. BIAS was smaller for the I(FP) compared to I(VM), and smallest in the I(VM+FP) condition. (**c**) Normalized BIAS across conditions for all subjects (N = 7) shown in different colors. Normalized BIAS was obtained by dividing BIAS in all conditions with BIAS in the I(VM+FP) condition. Across subjects, BIAS patterns were similar to the typical subject in panel b.
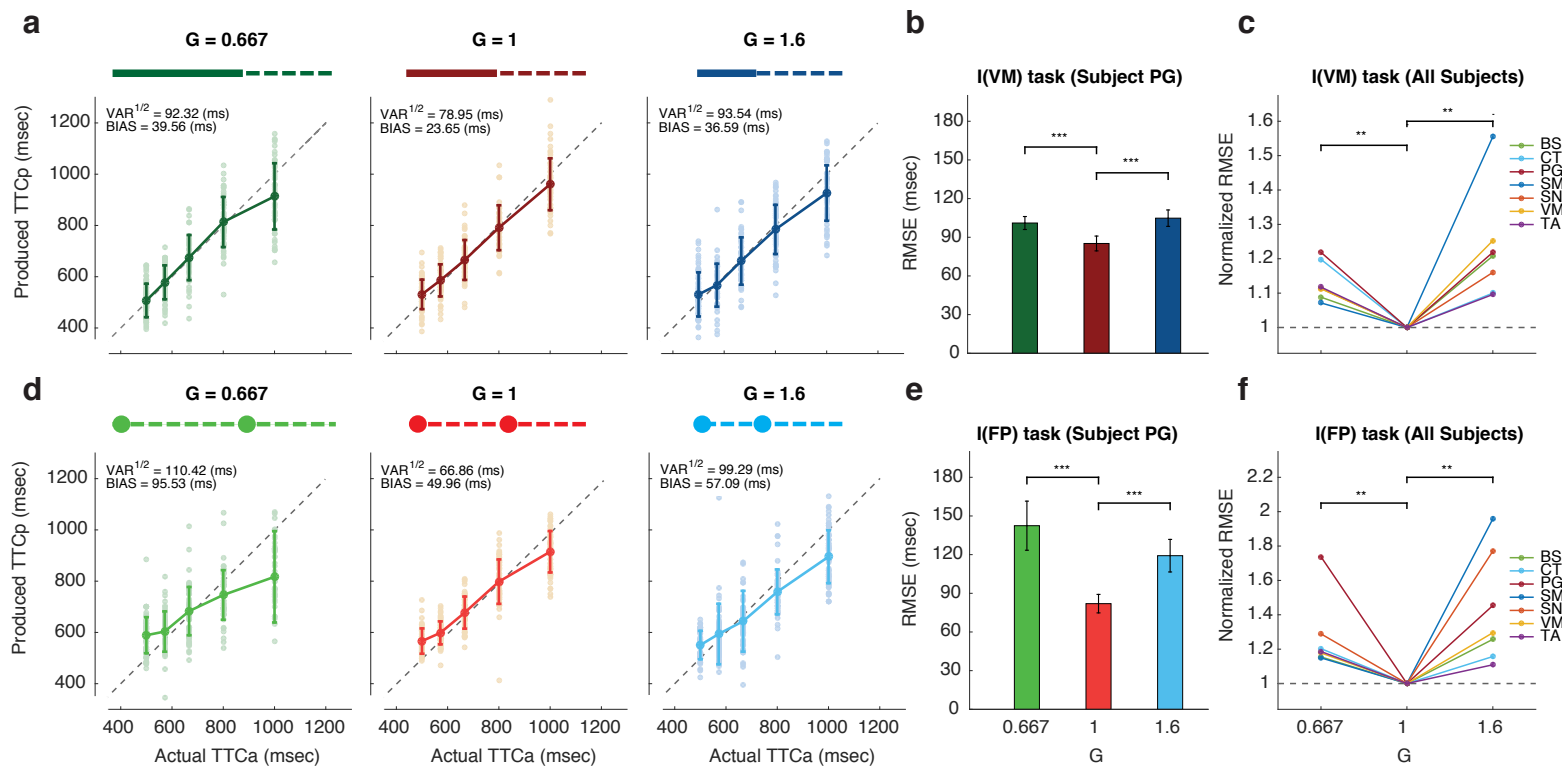
Figure 3. Interception using speed and implicit knowledge of temporal context (Experiment 2)

(**a**) Behavior of a typical subject in three variants of the I(VM) task with three different visible lengths and the same occluded length. Each condition was identified by a gain factor (G) that quantified the ratio of the occluded to visible length. Since the bar moved at a constant speed throughout each trial, the gain also reflected the ratio of the duration of occluded and visible sections of the path. Performance was quantified by comparing subjects' produced time-to-contact (*TTCp*) to the actual time-to-contact (*TTCa*). Light dots and dark circles show *TTCp* in each trial and the corresponding averages for each *TTCa*. BIAS was defined as described in Figure 2. VAR is the average variance of *TTCp* over the five intervals tested (*TTCa*) of the prior distribution. (**b**) Comparison of performance in terms of RMSE across conditions for a typical subject in I(VM) task. We estimated the standard error of RMSE through resampling data with 100 repetitions. (**c**) Normalized RMSE as a function of G for I(VM) task across all subjects (N = 7). RMSE in each condition was divided by RMSE when the gain was identical (G = 1). Different colors lines correspond to different subjects. (**d**) Behavior of a typical subject for I(FP) task with different gains (G) in Experiment 2. (**e**) Comparison across conditions for a typical subject in I(FP) task. (**f**) Normalized RMSE as a function of G for I(FP) task across all subjects (N = 7). For all subjects and in both tasks, RMSE was smallest for G=1 indicating best performance when the visible and occluded lengths were the same length (identity context, see main text).
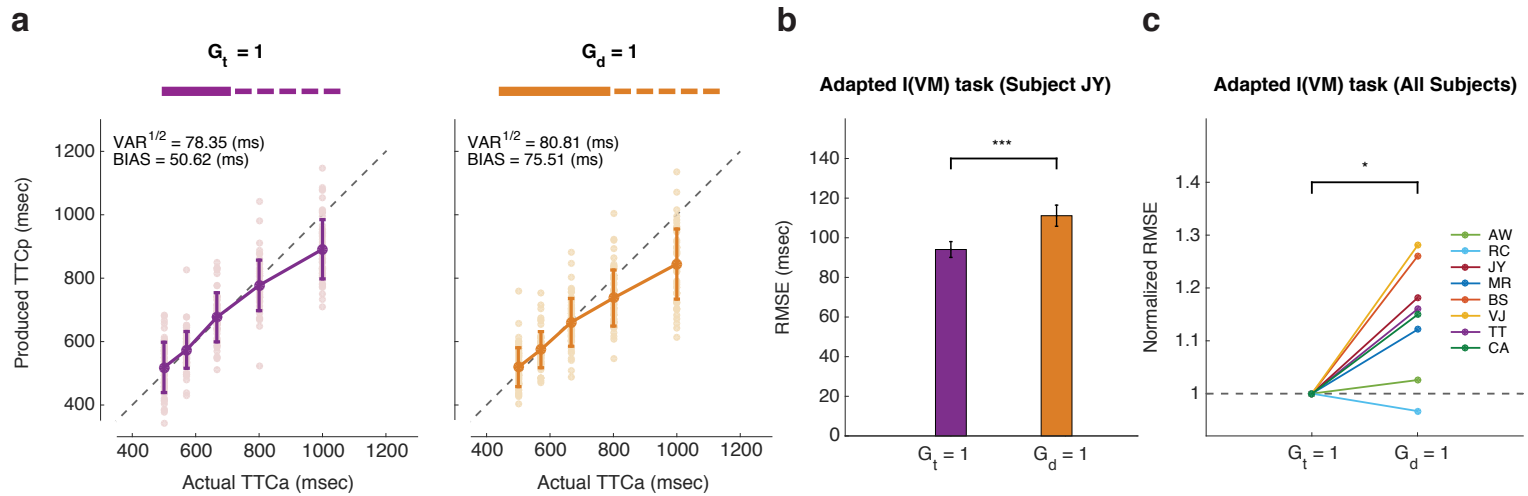
Figure 4. Interception using distance and temporal identity contexts (Experiment 3)

(**a**) Behavior of a typical subject for two variants of the I(VM) task, the temporal identity context ($G_t = 1$) and the distance identity context ($G_d = 1$). In both variants, unbeknownst to the subject, the speed behind the occluder was multiplied by 1.25 (25% faster than the visible section). $G_t = 1$: The durations of movement in the visible and occluded sections were the same. Because of speed difference between the two sections, the visible distance was shorter than the occluded distance. $G_d = 1$: The visible distance was same as the occluded distance, but the corresponding durations were different. (**b**) Comparison between the two conditions of I(VM) shown for a typical subject. We estimated the standard error of RMSE through resampling data with 100 repetitions. (**c**) Normalized RMSE across all subjects (N = 8). Different colored lines represented different subjects.
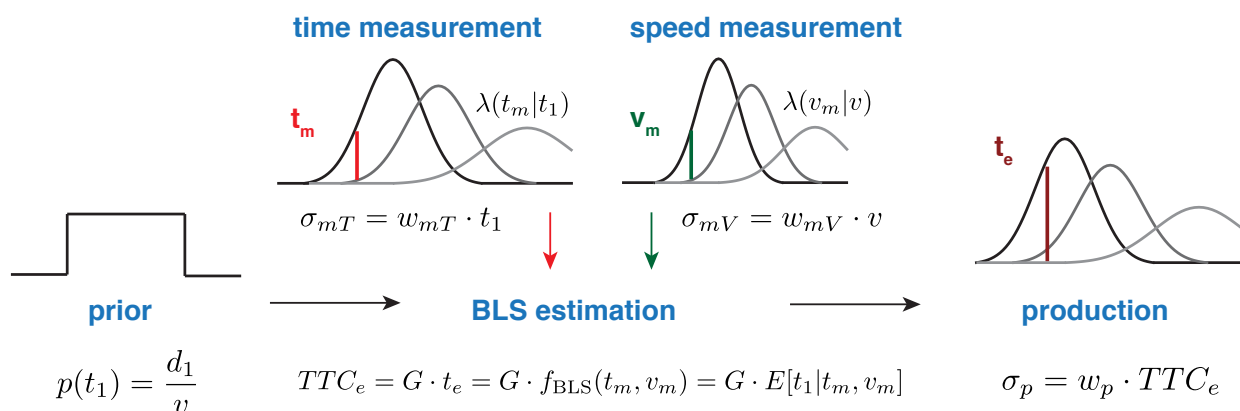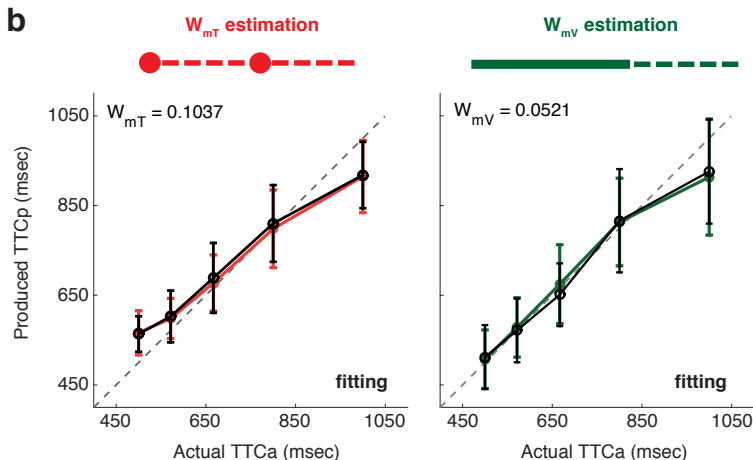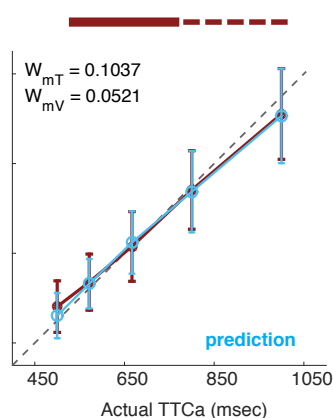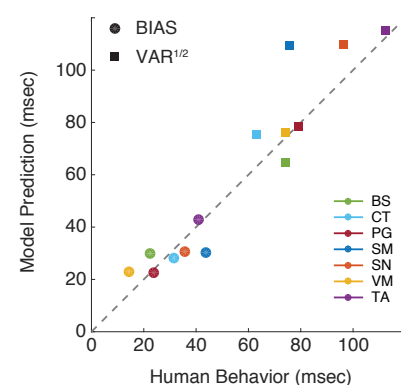
Figure 5. The Bayesian observer model of interception integrating speed and timing cues

(**a**) The Bayesian observer model for the I(VM) task. On each trial, the speed (v) was drawn from a uniform prior distribution. We used the relationship between the distance of the visible section (d_1) and speed to express the prior in terms of the duration the bar is visible ($p(t_1)$; left). We assumed that the observer makes two conditionally independent measurements of v and t_1, which we denoted by v_m (red vertical line) and t_m (green vertical line), respectively. We assumed that v_m and t_m are perturbed by zero-mean Gaussian noise with standard deviations (σ_{mV} and σ_{mT}) proportional to v and t_1 (top Gaussian curves) with constant of proportionality of w_{mV} and w_{mT}, respectively. The Bayesian observer computes the posterior from the likelihood functions, $\lambda(v_m|v)$ and $\lambda(t_m|t_1)$, and the prior, and uses a Bayes-Least-Squares (BLS) estimator, f_{BLS}, to infer the movement duration in the visible section, which we denoted by t_e (brown vertical line) from v_m and t_m. This estimate is then multiplied by the distance gain (G) to obtain an optimal estimate of time-to-contact ($TTCe$). Finally, the model incorporates motor variability via additional noise in the production stage. We modeled this noise as a sample from a zero-mean Gaussian with standard deviation scaling with TTC_e with scaling factor w_p. (**b**) The left panel (w_{mT} estimation) shows the behavior of a Bayesian observer model (red) fitted to the data (black) for a typical subject in I(FP) task with G = 1. Since the movement of the bar in the I(FP) task is not visible, we estimated w_{mT} from a Bayesian model that relies on the prior and t_m, but not v_m. The right panel (w_{mV} estimation) shows the Bayesian model (green lines) and the corresponding data (black) for the I(VM) task with G = 0.667. In the I(VM) condition, the observer has access to both speed and time. Therefore, we estimated w_{mV} from a Bayesian model that uses the prior, t_m and v_m with w_{mT} inferred from I(FP) with G = 0.667. (**c**) Behavior (black) and model prediction (blue) for a typical subject in the I(VM) task with G = 1. The prediction was made based on a Bayesian model whose w_{mT} and w_{mV} were derived in (**b**). (**d**) Comparison of summary statistics (BIAS and $VAR^{1/2}$) between human behavior (abscissa) and predictions from a Bayesian model (ordinate) across subjects (N = 7). BIAS and $VAR^{1/2}$ of the model was computed based on averages of 100 simulations of the Bayesian observer model. Different colors correspond to different subjects