Integration of speed and time for estimating time to contact

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

| Citation          | Chang, Chia-Jung, and Mehrdad Jazayeri. “Integration of Speed and Time for Estimating Time to Contact.” Proceedings of the National Academy of Sciences 115, no. 12 (March 5, 2018): E2879–E2887. |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| As Published      | http://dx.doi.org/10.1073/PNAS.1713316115                                                                                                                                               |
| Publisher         | Proceedings of the National Academy of Sciences                                                                                                                                          |
| Version           | Final published version                                                                                                                                                                  |
| Citable link      | http://hdl.handle.net/1721.1/119619                                                                                                                                                       |
| Terms of Use      | Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use. |
Integration of speed and time for estimating time to contact

Chia-Jung Chang and Mehrdad Jazayeri

*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and †McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139

To coordinate movements with events in a dynamic environment the brain has to anticipate when those events occur. A classic example is the estimation of time to contact (TTC), that is, when an object reaches a target. It is thought that TTC is estimated from kinematic variables. For example, a tennis player might use an estimate of distance \( d \) and speed \( v \) to estimate \( TTC = d/v \). However, the tennis player may instead estimate TTC as twice the time it takes for the ball to move from the serve line to the net line. This latter strategy does not rely on kinematics and instead computes TTC solely from temporal cues. Which of these two strategies do humans use to estimate TTC? Considering that both speed and time estimates are inherently uncertain and the ability of the human brain to combine different sources of information, we hypothesized that humans estimate TTC by integrating speed information with temporal cues. We evaluated this hypothesis systematically using psychophysics and Bayesian modeling. Results indicated that humans rely on both speed information and temporal cues and integrate them to optimize their TTC estimates when both cues are present. These findings suggest that the brain’s timing mechanisms are actively engaged when interacting with dynamic stimuli.

time-to-contact estimation | cue combination | temporal context | Bayesian model | sensorimotor coordination

Movements of our body and of objects around us create temporal events that demand our attention and command appropriate behavioral responses. For example, to catch a bouncing ball one must determine the moment the ball reaches the hand. To capture a tennis shot on camera one must anticipate the moment the ball reaches the racket. To catch an escaping prey the predator has to determine the time of the final leap. To shoot a flying disk one must estimate the moment to pull the trigger. Anticipating and reacting to such movement-related temporal events require an ability to estimate time to contact (TTC), that is, the time when a moving entity reaches a target location.

How does the brain estimate TTC? Early studies hypothesized that humans rely on variables derived from an object’s visual angle and its rate of expansion on the retina, of which the so-called tau is a classic example (1–3). Later, this proposal was deemed inadequate as it failed to capture many empirical observations (4–17). Most current theories posit that TTC estimation results from computations that rely on kinematic information (4–17). Specifically, it is assumed that the brain uses information about distance, speed, and acceleration to determine when an object reaches a designated target point. In this view, if we denote the distance by \( d \) and assume that the object moves with constant speed, \( v \), TTC can be derived as \( TTC = d/v \). This seems like a natural solution and matches our intuition of how to compute time from kinematic variables. However, the algorithms the brain uses for computing TTC need not match what is taught in physics classrooms. Here, we asked whether humans solely rely on kinematics (e.g., speed and distance) or if they additionally rely on temporal cues.

We use an example to demonstrate the potential relevance of temporal cues as an independent source of information for estimating TTC. Imagine catching an approaching bouncing ball. At first glance it may seem that TTC can be readily inferred from kinematic variables and equations without any need to explicitly estimate when the ball bounces. However, if estimates of speed and position are unreliable, for example when it is too dim to see the ball, one may infer TTC from the temporal structure of the sounds the ball makes upon bouncing off the ground. This example highlights a general and unresolved question in sensorimotor processing: When estimating TTC, do we rely solely on kinematic equations, or do we additionally rely on timing information that can be derived from positional information associated with moving objects (Fig. 1)?

Several decades of research in human psychophysics suggest that humans estimate behaviorally relevant variables by integrating information from multiple modalities (18–21). For example, to estimate the size of an object, humans optimally combine visual and tactile information (19), and to reach for an object, subjects combine uncertain spatial cues with prior expectation (18). With these considerations in mind, we hypothesized that humans estimate TTC by combining kinematic variables derived from visual information (e.g., speed) with estimates of elapsed time derived from the times when an object appears at different locations (e.g., the time it takes for an object to move from one point in space to another). However, testing this hypothesis is challenging because when an object moves between two points the brain can either directly estimate speed from visual motion (22–24), or it may infer speed from measuring the time it takes for the object to move between various locations along the movement path.

Significance

Existing theories suggest that reacting to dynamic stimuli is made possible by relying on internal estimates of kinematic variables. For example, to catch a bouncing ball the brain relies on the position and speed of the ball. However, when kinematic information is unreliable one may additionally rely on temporal cues. In the bouncing ball example, when visibility is low one may benefit from the temporal information provided by the sound of the bounces. Our work provides evidence that humans rely on such temporal cues and automatically integrate them with kinematic information to optimize their performance. This finding reveals a hitherto unappreciated role of the brain’s timing mechanisms in sensorimotor function.

Author contributions: C.-J.C. designed research; C.-J.C. performed research; C.-J.C. analyzed data; M.J. provided intellectual guidance and discussions on experiment design and data analysis; and C.-J.C. and M.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence should be addressed. Email: mjjaz@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713316115/-/DCSupplemental.

Published online March 5, 2018.
To investigate the complementary role of speed and timing mechanisms we designed a series of experiments in which we varied the temporal structure between visible and occluded segments of the path to systematically manipulate the reliability of the speed and the temporal information independently (Fig. 2). Consistent with our hypothesis, we found that subjects integrated temporal information that is available during both visible and occluded segments of the path with speed information that is only available during the visible segment to improve their estimate of TTC. To better understand the nature of the underlying computations we compared subjects’ behavior to that of an ideal Bayesian observer who optimally integrates speed and timing information. Similar to work in other sensorimotor domains (18–21, 25–38), the model was able to accurately capture subjects’ estimation strategy, indicating that humans efficiently integrate prior statistics with measurements of both speed and elapsed time. These results highlight a hitherto unappreciated function of the brain’s capacity to utilize time—indeed independent of speed—to inform sensorimotor function while interacting with dynamic stimuli.

Results
We first describe the general task design that we employed for all of the experiments (Fig. 2). On each trial, subjects held their gaze on a fixation point (FP) at the center of the screen (XFP = 0) throughout the trial and viewed one stimulus to the left of FP (XInit) and another to the right of FP (XTar). After a random delay, a bar began to move horizontally from XInit toward XTar with a fixed speed, v. In each trial, v was sampled from a discrete uniform prior distribution with five values ranging between 8° and 16°/s (Fig. 2D). Subjects had to press a button the moment the bar reached XTar.

We tested subjects in three conditions (Fig. 2A–C). In the first condition the bar was initially visible and then occluded. The visible segment extended from XInit to a transition point, denoted by XVis. The subsequent occluded segment extended from XVis to XTar. The distance of the visible and occluded segments were denoted by d1 and d2, respectively, and added up to the full length of the movement path (L). We refer to this condition as V-O (visible-occluded; Fig. 2A). In the second condition, the moving bar was occluded throughout the path but the position of the bar was flashed at XInit and at FP. We refer to this condition by O-F-O (occluded-flashed-occluded; Fig. 2B). In the third condition, the bar was initially visible and was additionally flashed when it reached FP. We refer to this condition by V-F-O (visible-flashed-occluded; Fig. 2C).

Exp. 1: TTC Estimation Benefits from Explicit Timing Cues. In the first experiment, the path was 16°, and XInit and XTar were located symmetrically at a distance of 8° to the left and right of FP. In this experiment, subjects were tested in all three conditions (i.e., V-O, O-F-O, and V-F-O). As described by the space of hypotheses in Fig. 1, in conditions in which the moving bar is visible (V-O and V-F-O) subjects could adopt one of two strategies to perform the task. First, they could decide when to press the button by relying on an estimate of the bar’s speed, v (that denotes subjective estimate), derived from visual motion. We refer to this as the speed strategy (Fig. 1, speed strategy). Alternatively, subjects could rely on timing information to perform the task. For example, in the V-O condition they could derive an estimate of the duration of the visible segment t1 (that denotes subjective estimate) and scale it by the ratio of the distance between the occluded and visible segments (d2/d1). We refer to this as the timing strategy (Fig. 1, timing strategy). This timing strategy can also be used in the V-F-O condition and is the only strategy that can be used in the O-F-O condition, in which no explicit visual cue about speed is present.

Our aim was to compare behavior in these conditions to assess whether subjects combine both strategies to compute TTC (Fig. 1, integration strategy). To quantify behavior, we compared the time it took for the bar to go from FP to XTar, which we refer to as actual TTC (TTCa = XTar/v) to the subjects’ produced TTC (TTCP). In the V-O and V-F-O conditions, in which the bar was flashed at FP, we defined TTCP as the interval between the flash at FP and button press. In the V-O condition, quantification of TTCP was less straightforward because the bar was not flashed at FP. In this condition we first measured the interval between XVis and button press and then scaled that interval by the ratio of XTar to the occluded distance (XTar/d2). This method of estimating TTCP for the V-O condition assumes that the internal estimate of the speed after XVis remains relatively stable.

As evident from the TTCP pattern for a typical subject (Fig. 3A), subjects were able to perform the task in all three conditions with different degrees of sensitivity. TTCP values were variable and exhibited a characteristic regression to the mean in which the average TTCP for each TTCa was systematically biased away from the identity line and toward the mean of the average TTXa (750 ms). As numerous previous experiments have demonstrated, this bias toward the mean is indicative of a Bayesian estimation strategy in which subjects reduce uncertainty associated with their sensory measurements (of speed and/or time) by using their knowledge about the prior statistics of TTXa (19, 21, 25–38). We quantified this regression using a BIAS term that quantifies the overall deviation from the identity line (Materials and Methods). When measurements are accurate, responses would be on average unbiased (i.e., near the identity line), and the corresponding BIAS would be small. However, when measurements...
are highly noisy we expect stronger regression to the means and larger BIAS values. In our dataset there was significant BIAS in all conditions. The magnitude of BIAS was significantly smaller in the V-F-O condition compared with both the V-O condition (t_{198} = 26.6435, P < 0.001, Hedges’ g = 3.7537) and the O-F-O condition (t_{198} = 27.4602, P < 0.001, Hedges’ g = 3.8687) (Fig. 3B). This reduction in BIAS was observed for all of the subjects and was significant across subjects (Wilcoxon one-side signed-rank test, statistics = 28, P < 0.01) (Fig. 3C), suggesting that humans combine information gleaned from the visual motion

![Fig. 2. Behavioral task conditions. (A) V-O condition. (Top) Schematic illustration of the V-O condition in which a bar moves along a path that is divided to two segments, a visible segment extending from an initial point (X_{init}) to a transition point (X_{tran}) and an occluded segment (i.e., invisible) from X_{tran} to a target point (X_{tar}). (Bottom) Trial structure for the V-O condition. Subjects were asked to fixate at the central fixation point (gray circle). Afterward, a bar moved from X_{init} to the left of the fixation point (white circle) to X_{tran} to the right of the fixation point (white vertical line). The bar was visible initially and occluded afterward (Top). Subjects had to press a key when they judged the moving bar to have arrived at X_{tran}. When responses were sufficiently accurate (Materials and Methods), the moving bar and the target bar both turned green (shown). Otherwise, they both turned red (not shown). (B) O-F-O condition. (Top) Schematic illustration of the O-F-O condition in which the bar was only flashed at X_{init} and later at a position along the path (black circle). In experiments where we tested the O-F-O condition the distance between the two flashes was varied. (Bottom) Trial structure for the O-F-O condition in the same format as A. The example trial shows a case where the flashed position coincided with the fixed point, corresponding to the design in Exp. 1. In this example the feedback is shown as red, indicating a hypothetical trial where the response was too early. (C) V-F-O condition. This condition includes both an initial visible segment (from X_{init} to X_{tran}) and a flashed position some time after the visible segment (black circle). In experiments where we tested the V-F-O condition the flashed position was always in the middle of the segment. (Bottom) Trial structure for the V-F-O condition in the same format as A. (D) Prior distribution of the bar speed (v). Speed was fixed in each trial but was sampled from a discrete uniform distribution ranging between 8–16˚/s across trials.

![Fig. 3. TTC estimation using speed and explicit timing cues. (A) Behavior of a typical subject for different conditions in Exp. 1. (Left) Corresponds to the V-O condition (blue) where the bar was initially visible. (Middle) Corresponds to the O-F-O condition (red) in which the bar was flashed at the starting point and the central fixation point. (Right) Corresponds to the V-F-O condition (gray) with an initial visible segment and a flash at the fixation point. Performance was quantified by comparing TTCp to 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, respectively. The inset in each panel reports the overall BIAS in TTCp. BIAS was quantified as the average error over the five distinct TTCa (i.e., the root mean square of differences between five solid dark circles and the corresponding diagonal dash line on the plot) (Materials and Methods). (B) BIAS comparison across task conditions for a typical subject. We estimated the SE through resampling data with 100 repetitions. BIAS was smaller for the O-F-O compared with V-O and smallest in the V-F-O condition. See main text for statistics. **P < 0.001. (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 V-F-O condition. Across subjects, BIAS patterns were similar to the typical subject in B. See main text for statistics. **P < 0.01.
(speed and possibly timing) with the additional explicit timing cue provided by the flash at FP to reduce uncertainty.

Although BIAS provides an overall estimate of deviations from veridical TTCa, it does not specify the direction of bias. In other words, both positive and negative biases would lead to an overall increase in BIAS. To ensure that the direction of bias in the data was consistent with a regression toward the mean (i.e., overestimation of small TTCa and underestimation of large TTCa), we additionally quantified the relationship between TTCp and TTCa using linear regression. In all conditions the slope of the regression was significantly smaller than unity, indicating that the BIAS was indeed consistent with the hypothesized regression to the mean (Fig. S4).

While this result is consistent with subjects integrating the two cues, it is also possible that the flashed stimulus at FP was not used as an explicit timing cue, and instead was used to simply reset the internal estimate of the position of the bar along the path. To test this possibility, we tested a subset of subjects in a cue conflict version of the V-F-O condition in which the flash at FP occurred either at the correct time (i.e., when the bar reached FP), 100 ms too early, or 100 ms too late (Fig. S1). This experimental manipulation enabled us to distinguish between multiple hypotheses.

**H1: Position-reset hypothesis.** According to this hypothesis, at the time of flash subjects reset the position of the bar to the location of FP without changing their estimate of the speed of the bar and without using the time of the flash as an additional cue. Since we quantified TTCp from the time of the flash to the button press, we should see no change in the relationship between TTCp and TTCa. Note that this hypothesis was not explicitly considered in Fig. 1.

**H2: Speed-only hypothesis.** According to this hypothesis, subjects ignore the flashed stimuli entirely and therefore TTCp measured with respect to the time of flash would be shifted by the same amount as the jitter but in the opposite direction. In other words, TTCp should increase by 100 ms when the flash was presented 100 ms too early and decrease by the same amount when the flash was presented 100 ms too late. The original experiment already rejects this hypothesis since there were clear differences between subjects’ performance in the V-O and V-F-O conditions. However, the results from the jitter experiments could further validate the importance of the external timing cue. This hypothesis is referred to as the speed strategy in Fig. 1.

**H3: Timing-only hypothesis.** According to this hypothesis, subjects only rely on the timing cues and ignore the opportunity to estimate speed from the visual segment of the path. If subjects were only using the time of the flash to estimate TTC, the average TTCp should be shifted exactly by the same duration as the jitter and in the same direction. Therefore, TTCp should decrease by 100 ms when the flash was presented 100 ms too early and increase by the same amount when the flash was presented 100 ms too late. Again, the original experiment already rejects this hypothesis since there were clear differences between performance in the V-F-O and V-O conditions. However, we expected the results of the cue conflict experiments to also reject this hypothesis. This hypothesis is referred to as the timing strategy in Fig. 1.

**H4: Cue-integration hypothesis.** According to this hypothesis, the jittered flash time would alter the timing-based evidence and would therefore cause a concomitant bias in TTCp. Importantly, however, this bias should be less than the size of the jitter (i.e., less than 100 ms) since temporal cues only serve as part of the information that guides subjects’ behavior (the other part being the speed information gleaned from the visual segment of the path). This hypothesis is referred to as the integration strategy in Fig. 1.

We found that TTCp changed significantly in the presence of jittered flashes (t test, P < 0.001 for subjects JT, CN, and BS and P = 0.377 for subject MD), which rejected H1 and H2, and that the overall shift in TTCp was significantly smaller than 100 ms (t test, P < 0.001 for all subjects), rejecting H3 (Fig. S1). Together with the main results of Exp. 1, the observations indicate that subjects integrated both speed and explicit timing information to estimate TTC.

**Exp. 2: TTC Estimation Benefits from Implicit Timing Cues.** The performance improvement in V-F-O compared with V-O and O-F-O demonstrated that humans benefited from an explicit timing cue provided by the flash at FP. This raises the intriguing possibility that humans utilize timing information even if it is not in the form of an explicit flashed position. For example, it may be the case that even in the V-O condition where there are no flashed stimuli subjects determine when to press the button by combining two cues, one coming from speed information (e.g., Xプ/υ) and the other from scaling the duration of the visible segment (t1) by the ratio of the occluded to visible segments (d2/d1). The former follows directly from kinematic equations (e.g., “if speed is doubled, it would take half as long”), and the latter derives from an ability to scale time intervals (e.g., “if distance is doubled, it should take twice as long”).

To validate if such implicit temporal cue is used for estimating TTC, we designed a variant of the V-O condition in which we varied the ratio of the visible and occluded segments (d1 and d2, respectively). We reasoned that when the visible and occluded intervals have the same duration, subjects would find the timing information more reliable and give it more weight for estimating TTC. However, since our objective was to assess the importance of implicit timing, we needed to make sure that varying the visible segments would not cause an appreciable change in the subjects’ estimate of speed. Therefore, we first designed an experiment to quantify how changes in the visible segment influence the accuracy of subjects’ speed estimate (Fig. S2).

The experiment was a variant of the V-O condition in which we changed the length of the visible segment parametrically from 0.625° to 5° in log scale while keeping the occluded distance fixed at 8° (Xプ = 8°). We found that performance quantified in terms root-mean-squared error (RMSE) improved significantly as the visible lengths increased from 0.625° to 1.25° (paired-sample t test, t399 = 56.61, P < 0.001) and saturated afterward (paired-sample t test, t399 = 0.9031, P = 0.3670). In other words, the fidelity of the speed estimate saturated at a visible length of 1.25°.

We then tested subjects’ behavior in the V-O condition in a separate experiment where we varied the ratio of the visible and occluded segments while keeping the visible length longer than the empirically observed saturation point of 1.25°. This ensured that any change in performance could not be attributed to an improvement or degradation of speed estimates and must therefore reflect a capacity to measure and scale time intervals. We tested subjects’ performance in three conditions. In all conditions the occluded length was fixed (d2 = 8°). Across conditions the ratio of the occluded to visible length (d2/d1) was varied by a gain factor (G = d1/d2). The three gain factors were 0.667, 1, and 1.6.

Fig. 4A and B show the performance of a typical subject in the three conditions. Notably, the best performance was not associated with G = 0.667 when the visible length was longest. Instead, RMSE was smallest when the visible and occluded lengths were equal (G = 1.6, t198 = 20.3981, P < 0.001, Hedges’ g = 2.9308, G = 1.6, t198 = 22.9261, P < 0.001, Hedges’ g = 3.2299), which we refer to as the temporal identity condition. The same was true across subjects (Fig. 4C, Wilcoxon one-sided signed-rank test, statistics = 28, P < 0.01), revealing a systematic and consistent improvement of performance in the identity condition (i.e., when G = 1).

The same subjects were also tested in the O-F-O condition, and for the same three gain values. As evident from the behavior of the same typical subject, RMSE was smaller when the two segments (i.e., before and after the flash) were the same (Fig. 4D and E) compared with when the first segment before the flash was longer (G = 0.667, t198 = 29.7316, P < 0.001, Hedges’ g = 4.1887) or shorter (G = 1.6, t198 = 25.6390, P < 0.001, Hedges’
g = 3.6122). This effect was present across subjects (Fig. 4F; Wilcoxon one-side signed-rank test, statistics = 28, P < 0.01), indicating that temporal identity helped subjects improve their estimate of TTC.

One potential concern in the case of G = 1 is that subjects may have detected that the two segments were temporally identical and switched to a purely timing strategy. To evaluate this possibility we compared subjects’ performance in the V-O and O-F-O conditions in the specific case when the two segments are identical (Fig. S3). If subjects were only relying on a timing strategy, we would expect subjects’ performance in these two conditions to be the same (since both have the timing information with G = 1). However, if the presence of G = 1 only serves to make the timing cue more reliable, we would expect performance to be better in the V-O condition since that condition provides the additional speed-dependent information.

We found that RMSE was consistently and significantly smaller in the V-O condition (Wilcoxon one-side signed-rank test, statistics = 3, P < 0.001), ruling out the hypothesis that G = 1 motivated subjects to abandon the speed information and rely only on the identity temporal structure. These results suggest that subjects exploited the temporal structure in addition to speed cue to improve their performance.

One question that we did not address in this experiment is why G = 1 provides a more reliable timing cue. While it is not surprising that reproducing a time interval may be more reliable than producing an interval that is scaled by an arbitrary gain factor, a detailed quantification of this factor requires additional experiments. However, our results are fully consistent with a recent study (39) that demonstrated that mental transformation of time intervals (i.e., multiplication by a gain) increases noise levels and reduces the reliability of timing cues.

Exp. 3: TTC Estimation Improves with Temporal Identity. Exp. 2 clearly demonstrated that TTC estimation was most accurate when the visible and occluded segments of the path were identical. This is consistent with our hypothesis that performance benefited from temporal identity (i.e., G = 1). However, it is also possible that when the visible and occluded segments are the same length (d_1 = d_2) performance improves because subjects can estimate the distances more accurately. In other words, it may be that subjects benefited from distance identity (same lengths) and not temporal identity (same durations). This seems unlikely given that d_2 was fixed throughout all experiments. Nonetheless, we conducted an additional experiment to assess which of the two properties helped subjects in estimating TTC.

Since distance and duration are related through speed, the only way to dissociate the two is to make the speed of the bar differ between the visible and occluded parts of the path. Therefore, we designed a variant of the V-O condition in which, unbeknownst to the subjects, the speed of the bar behind the occluder was made 1.25 times faster than the speed in the visible portion (Fig. 5A). The nonidentical speed ratio enables us to create conditions in which the distance and temporal identity were dissociated. In one condition, the visible and occluded distances were the same but the corresponding durations were not. We refer to this condition as the distance identity condition (G_d = 1, G_s = 0.8). In another condition, we matched the ratio of the distances to the ratio of the speeds so that the corresponding...
durations were the same. We refer to this condition as the
temporal identity condition \( G_x = 1 \), \( G_d = 1.25 \).

New subjects were recruited for this experiment to ensure that
sensitivity to temporal identity could not be attributed to
exposure to previous experiments. Since subjects were not aware of
the speed change behind the occluder, they could only adjust
their performance based on feedback. We compared subjects’
performance between the \( G_d = 1 \) and \( G_x = 1 \) conditions. We
reasoned that an observer that relies on the distance identity
should have better performance (lower RMSE) in the \( G_d = 1 \)
condition. In contrast, an observer that relies on the temporal
identity would have a lower RMSE in the \( G_x = 1 \) condition de-
spite the fact that the distances between the visible and occluded
parts are not the same.

We found that RMSE was lower for the temporal identity
compared with the distance identity condition as shown for a typical
subject (Fig. 5 B and C; \( t_{198} = 25.6431, P < 0.001 \), Hedges’ \( g =
3.6127 \)) and across subjects (Wilcoxon one-side signed-rank test,
statistics = 34, \( P < 0.05 \)). This finding further substantiates our
conclusion that subjects benefit from temporal identity. We note
that this experiment does not rule out a potential complementary
role for distance identity, but it reveals the importance of tem-
poral structure in estimating TTC.

**Bayesian Integration of Speed and Time Explains Performance of TTC**

Estimation. Exps. 1–3 established that subjects rely on both speed
and timing strategies to estimate TTC. Another salient feature of
subjects’ behavior across all conditions was the regression to the
mean of \( TTCp \) across the range of \( TTCa \). This was true for the
external timing cue conditions in Exp. 1, for the implicit timing
condition in Exp. 2, and in the control condition in Exp. 3. This
observation suggests that, in addition to speed and timing infor-
mation, subjects rely on their prior knowledge of the range of
TTCa they encounter in the experiment. Following previous
work \( (38) \), we hypothesized that the subjects’ responses may
follow the prediction of a Bayesian model that optimally inte-
grates both the speed and timing measurements with the prior
distribution (Materials and Methods) (Fig. 6d).

To test this hypothesis rigorously we developed a Bayesian
observer model to explain subjects’ behavior in the \( V-O \) condi-
tion. The observer model combined two conditionally independent
measurements from the visible segment of the path, one

\[
\text{ associated with the speed of the bar (} v_m \text{) and another
associated with the duration of the first visible segment (} t_{1m} \text{).}
\]

Following previous work, we assumed that these measurements were sub-
ject to scalar variability \( (28, 40–43) \). In particular, we assumed
that the SD of noise on speed scaled with the bar’s speed \( v \) with
constant of proportionality \( w_{mv} \) and SD of noise on elapsed
time scaled with visible duration \( t \) with constant of propor-
tionality \( w_mT \). The ideal observer first computes the pos-
terior from the product of the prior, \( p(t) \), the likelihood of the
bar speed, \( \lambda(t_{1m} | v) \), and the likelihood of the visible duration,
\( \lambda(t_{1m} | t) \), and then uses the mean of the posterior as the optimal
estimate of TTC (\( TTC_e \) in Fig. 6). In other words, \( TTC_e \) corre-
sponds to the so-called Bayes-least-squares (BLS) estimate. To
compare the model to subjects’ behavior we augmented the ideal
observer with a production stage by adding scalar noise to the
\( TTC_e \) with constant of proportionality \( w_p \) to generate \( TTCp \).

We first estimated \( w_{mv} \) and \( w_{mT} \) for each subject. In most
Bayesian models the model is evaluated by assessing the goodness
of fit. A more powerful approach is to fit the model to a training
dataset and examine how well it explains a test dataset. An even
more powerful approach is to fit the model to one set of condi-
tions and ask whether it predicts data in another condition to
which it was not fitted. We employed the last approach. For each
subject we estimated \( w_{mp} \) from the \( O-F-O \) condition with \( G = 1 \)
(Fig. 6B, Left), and \( w_{mp} \) from the \( V-O \) condition with \( G = 0.667 \)
(Fig. 6B, Right) and used those estimates to predict subjects’ be-
behavior in the \( V-O \) condition with \( G = 1 \) (Fig. 6C).

To estimate \( w_{mp} \) we developed a Bayesian observer for the
\( O-F-O \) condition with \( G = 1 \). In this case, the sensory information
provided was the interval between when the bar flashed at initial
start and when it reached FP (halfway along the path), which we
denote by \( t_f \). We fitted subjects’ behavior by a BLS estimator that
only relied on the likelihood of \( t_f \), \( \lambda(t_{1m} | t_f) \) and the prior distri-
bution, \( p(t_f) \). As shown for one subject (Fig. 6B, Left), and
consistent with previous work in a similar task \( (37, 44–46) \), the
model accurately captured behavior.

Next, we estimated \( w_{mp} \) from fits of the Bayesian model to the
\( V-O \) condition with \( G = 0.667 \). For this fitting procedure we used
the corresponding \( w_{mp} \) derived from the \( O-F-O \) condition with
\( G = 0.667 \) (Materials and Methods). As shown for the same
subject (Fig. 6B, Right), the model successfully accounted for the
behavior. Recall that in the \( V-O \) condition we had made the

![Fig. 5. TTC estimation in the distance identity and temporal identity conditions. (A) Behavior of a typical subject for two variants of the V-O condition (Exp. 3), the temporal identity condition \( (G_x = 1 \), that is, same visible and occluded durations) and the distance identity condition \( (G_d = 1 \), that is, same visible and occluded distances). In both variants, unbeknownst to the subject, the speed in the occluded segment was multiplied by 1.25 (25% faster than the visible section). \( G_x = 1 \) (purple): 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 longer than the occluded distance (i.e., \( G_d = 0.8 \)). \( G_d = 1 \) (orange): The visible distance was same as the occluded distance, but the corresponding durations were different (i.e., \( G_x = 1.25 \)). (B) Comparison between these two conditions for a typical subject. We estimated the SE of RMSE through resampling data with 100 repetitions. See main text for statistics. ***P < 0.001. (C) Normalized RMSE across all subjects \((n = 8) \). Different colored lines represented different subjects. See main text for statistics. *P < 0.05.](image-url)
visible length long enough so that subjects’ estimate of speed had saturated and was thus no longer dependent on \( G \) (Fig. S2). This allowed us to safely use the fit to \( \hat{w}_{mT} \) derived from the \( G = 0.667 \) condition to predict behavior in the \( G = 1 \) condition.

Finally, we used each subject’s fits to \( \hat{w}_{mV} \) and \( \hat{w}_{mT} \) to predict the behavior in the \( V-O \) condition for \( G = 1 \). The model was able to predict the observed \( \text{TTC}_{p} \) values as shown for one example subject (Fig. 6C) and captured the summary statistics (BIAS and VAR) across subjects (Fig. 6D). This is remarkable considering that both \( \hat{w}_{mV} \) and \( \hat{w}_{mT} \) were estimated from other task conditions and provides strong support that subjects integrate prior information, speed information, and timing information to optimize their estimate of TTC.

A point of potential concern in our modeling work is that we modeled both the prior distribution over the speed and the time intervals as uniform. This formulation is inaccurate. Given that the objective prior distribution of speed was uniform and that duration was inversely proportional to speed, the objective prior on intervals cannot be uniform. In our original model, we made this approximation because both distributions were discrete and because the exact formulation of the prior was not relevant to our main conclusion about the integration of the likelihood functions associated with speed and duration. However, to ensure that our results did not depend on the specific assumption of a uniform prior over time intervals, we constructed another model in which the prior more accurately reflected the distribution used in the experiment. For this model, we followed previous work (47) and derived a “subjective” prior of time intervals by blurring (i.e., convolving) the objective distribution with a normal distribution whose SD was proportional to elapsed time. This alternative formulation did not change our main conclusion about the integration of speed and timing information but provided an overall better fit to behavior, suggesting that subjects relied on the empirically observed samples to form their prior over time intervals (Fig. S5).

Discussion

Current models assume that estimation of TTC between the body and an object or between two objects depends on measurements of kinematic variables such as speed, distance, and/or
depth (4–17, 48–51). Our work reveals that humans additionally exploit timing information gleaned from the temporal structure of events in the environment as an alternative source of information to estimate TTC. Moreover, we show that humans automatically combine this timing strategy with kinematics to derive more accurate estimates of TTC.

We demonstrated the role of timing in two complementary sets of experiments. In the first set of experiments we presented subjects with a task in which estimation of TTC could benefit from either timing or speed information. Results indicated that when explicit timing cues were available subjects integrated timing information with their measurements of speed to derive more accurate estimates of when a moving bar would reach a target position. This result extends a large body of evidence showing that humans fuse information from multiple modalities to improve their performance (19–21).

In the second set of experiments we removed the explicit timing cue and instead asked whether subjects could exploit implicit timing cues in the environment. In our experiment we varied the ratio between the intervals when the bar was visible and occluded. Based on recent work (39), we hypothesized that when the visible and occluded segments have the same duration subjects would automatically make use of this temporal identity and rely more on the timing information to estimate TTC. Results validated that subjects relied on the temporal identity structure to improve their performance. Notably, performance in the temporal identity context was even better than when the occluded length was the same and the visible length was made longer. In other words, prolonging the visible portion, which could only improve subjects’ estimate of speed, was harmful to performance when it broke the temporal structure conferred by the identity context. This result powerfully demonstrated that the key factor driving the performance improvement was the presence of temporal identity. This conclusion was reinforced by a control experiment showing that the result was due to temporal—not distance—identity. Finally, it is also important to note that the role of timing strategy in our experiment cannot be attributed to apparent motion because the distances and time intervals we used in our experiment were well outside the range that typically induces an apparent motion percept (52).

Our work also intersects with the body of work revealing subjects’ ability to integrate sensory information with prior expectations (19, 21, 25–38). This integration is often characterized in the context of Bayesian models that formalize how prior knowledge and sensory cues must be integrated to optimize performance. We found that a Bayesian model that optimally integrates the prior distribution of TTC with evidence derived from both speed and temporal cues accurately captured subjects’ behavior. This result suggests that the human brain is optimized to combine speed and time information for object interception. Note that the integration of speed and time information is distinct from the indirect role that time would play by improving one’s estimate of speed (53–55). As we demonstrated in the control experiment (Fig. S2), the improvement of speed estimate saturates rapidly as viewing time increases and cannot account for our finding in Exp. 2. In other words, our results reveal that humans actively integrate elapsed time with speed information to estimate TTC.

These experiments lead to a simple conclusion that humans actively engage timing mechanisms during estimation of TTC. To put this finding in context, it is important to distinguish between the role of time during the visible and occluded regions of the path. When an object moves behind an occluder, subjects can no longer measure the object’s speed and thus have no choice but to rely on their sense of time. This idea was formalized by Treysillan and others in relation to humans’ ability to extrapolate an object’s location behind an occluder (56, 57). This is fundamentally different from what we propose; our findings indicate that humans actively integrate information about temporal contexts and events even when the object is visible. In other words, timing seems to be an integral component of how we interact with dynamic stimuli, both to better estimate where they are when they are visible and to infer where they might be when they are occluded.

One important implication of our work is for studies of object interception. Real-world object interception involves a decision to initiate a movement followed by ongoing adjustments based on sensorimotor feedback. Although successful interception requires a tight coordination between the initiation and the subsequent adjustments, the two processes typically involve different computations (58). The decision of when to initiate is determined by a prediction of how long it would take to reach the object, which is directly related to our work on TTC estimation. While our work does not address any potential role of timing for the sensorimotor coordination after movement initiation, it does invite a revision of the computational models that specify how the brain computes the movement initiation time. In particular, it suggests that the cognitive and/or motor planning stage of interception behavior may be particularly sensitive to preceding temporal events in the environment, as recent physiology experiments suggest (44). It is also consistent with numerous imaging and electrophysiological studies that find an important role for premotor and supplementary motor areas in timing (59–63). In contrast, temporal cues may not play an active role during the adjustments that follow movement initiation when the brain has access to movement-related, state-dependent information (64–66).

It is worthwhile considering why the role of time was not noted in prior research on object interception. We think that the answer has to do with the simplicity of behavioral tasks commonly used in laboratory experiments (but see refs. 67 and 68). Many previous experiments lacked a rich spatiotemporal context that could reveal the relevance of temporal structure. However, real-world examples of object interception take place in the presence of temporal statistics, spatial landmarks, and temporal events such as collisions and/or reflections, all of which make knowledge about time highly informative. A notable observation in our experiment was that TTC estimates were more accurate in the temporal identity context, which replicates results from a recent study showing nonidentity transformations are associated with higher sensorimotor noise (39).

We speculate that the improvement of performance we found in the temporal identity context may be an instance of a more general principle related to temporal structures for which the human brain has a strong internal prior. If so, we would expect stronger effects of timing information in the presence of sounds that form rhythms or for integer ratios for which strong internal priors have been reported (69). A real-world example of this conjecture applies to intercepting a bouncing ball. According to our results, we predict that subjects benefit from the bounce sound, especially when visual information is uncertain (e.g., dribbling a basketball without looking at the ball). These considerations highlight the need for future research to move beyond simple behavioral tasks and examine object interception and TTC estimation in more naturalistic settings where the underlying dynamics are governed by richer spatiotemporal contexts. Exploration of behavior in more naturalistic settings may further substantiate the importance of temporal events and contexts in processing dynamic stimuli.

**Materials and Methods**

All experiments were approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology, and all subjects provided informed consent before participation. We used three experiments to examine how people infer TTC. Seven adult subjects participated in Exp. 1. A different group of seven adult subjects participated in Exp. 2. Another group of eight adult subjects participated in Exp. 3. All subjects had normal or corrected-to-normal vision. In all experiments we quantified behavior by comparing the statistics of experimentally specified actual TTC (TTCa) with the subjects’ TTC (TTCp).

We developed and fitted Bayesian observer model to describe performance in the V-O condition (Fig. 6A) based on previous work on interval
reproduction (38). Model details and fitting procedures are provided in Supporting Information.

Instead of fitting the Bayesian model to each dataset, we asked whether we could fit the model to some conditions and then use parameters of the fitted model to predict behavior in other conditions. We aimed to predict behavior in the most important condition where subjects integrated speed with the identity temporal context (i.e., V-O with G = 1). We assumed that the noise associated with the measurement of time is the same in the V-O and O-F-O conditions and therefore used the Bayesian model to the O-F-O condition for G = 1 to estimate \( \text{w}_{\text{M}} \) (Fig. 6B, Left). We further assumed that the measurement of speed in V-O condition would be the same across two different gains (G = 1 and G = 0.667), given that the accuracy of speed measurement saturated rapidly (Fig. 52). We first found \( \text{w}_{\text{M}} \) for G = 0.667 from the O-F-O condition and then used this value to fit a Bayesian model to the V-O condition with G = 0.667 to estimate \( \text{w}_{\text{M}} \) (Fig. 6B, Right). Finally, we used the \( \text{w}_{\text{M}} \) inferred from O-F-O with G = 1 and \( \text{w}_{\text{M}} \) inferred from V-O with G = 0.667 to predict behavior in the V-O condition with G = 1 (Fig. 6C).