Mouse cursor trajectories capture the flexible adaptivity of predictive sentence processing

Anuenue Kukona[1] and Nabil Hasshim[2]

[1] School of Human Sciences, University of Greenwich

[2] School of Applied Social Sciences, De Montfort University

Author Note

Anuenue Kukona https://orcid.org/0000-0003-4377-3057

Nabil Hasshim https://orcid.org/0000-0001-6375-8554

Nabil Hasshim is now at the School of Health and Society, University of Salford.

Data collection was supported by funding from the Institute for Psychological Science, De Montfort University.

This study was not preregistered.

All data and code have been made publicly available at OSF and can be accessed at https://osf.io/ndsmf

Correspondence concerning this article should be directed to Anuenue Kukona, University of Greenwich, Old Royal Naval College, Park Row, London SE10 9LS, UK.

Email: a.p.bakerkukona@greenwich.ac.uk

Abstract

Recent psycholinguistic findings raise fundamental questions about comprehenders' ability to rationally adapt their predictions during sentence processing. Two mouse cursor tracking experiments (each *N* = 85) assessed this adaptivity by manipulating the reliability of verb-based semantic cues. In Experiment 1, predictive mouse cursor movements to targets (e.g., bike) vs. distractors (e.g., kite) were measured while participants heard equal proportions of non-predictive (e.g., "spot … the bike"), predictive (e.g., "ride … the bike") and anti-predictive (e.g., "fly … the bike") sentences. In Experiment 2, participants heard equal proportions of non-predictive and anti-predictive sentences. Participants were observed to flexibly adapt their predictions, such that they disengaged prediction in Experiment 1 when verb-based cues were unreliable and as likely to be disconfirmed as confirmed, while they generated adapted predictions in Experiment 2 when verb-based cues were reliably disconfirmed. However, links to individual differences in cognitive control were not observed. These results are interpreted as supporting rational theoretical approaches.

Keywords: Adaptation: Mouse cursor tracking: Prediction: Rationality; Sentence processing

Mouse cursor trajectories capture the flexible adaptivity of predictive sentence processing

Prediction is an important focus of the sentence processing literature (e.g., see reviews by Altmann & Mirković, 2009; Federmeier, 2007; Huettig, 2015; Kamide, 2008; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Van Petten & Luka, 2012). However, recent findings (e.g., Van Wonderen & Nieuwland, 2023; Zhang et al., 2019) highlight the surprising inflexibility (i.e., non-adaptability) of predictive sentence processing. The aim of the current research was to both assess comprehenders' ability to adapt their predictions and capture this adaptivity using mouse cursor tracking.

Adaptation is widely documented in the sentence processing literature. For example, Wells et al. (2009) focused on object relative clauses (e.g., "The clerk that the typist trained told the truth about the missing files."), which are both atypical (e.g., the object "the clerk" preceding rather than following "the typist trained…" contrasts with the typical subject-verb-object ordering of English) and typically difficult to process (e.g., see Gibson, 1998). However, they found that this difficulty was reduced when participants were exposed to relative clauses over multiple study sessions. These findings suggest that comprehenders may adapt their syntactic expectations based on experience to facilitate processing. Related evidence for syntactic adaptation suggests that participants can also adapt their expectations within a study session (e.g., Fine et al., 2013), link their expectations to individual talkers (e.g., Kamide, 2012) and adapt to a range of structures (e.g., also see coordination; Kaan et al., 2019; Dempsey et al., 2024).

Evidence for adaptation supports rational theoretical approaches. Accordingly, comprehenders are assumed to use their probabilistic knowledge of language (i.e., rationally) to maximise successful comprehension. For example, Levy (2008) hypothesises that at each word in an unfolding sentence, comprehenders generate beliefs about the underlying syntactic structure based on their knowledge, which they (i.e., rationally) update using Bayes' rule. Kuperberg and Jaeger (2016) emphasise that this theoretical approach is "inherently predictive" because beliefs generated at each word correspond to probabilistic predictions about the next word(s) (e.g., Bayesian posteriors become priors with each cycle of belief updating). Crucially, findings like Wells et al. (2009) suggest that comprehenders can adapt their knowledge to reflect changing language statistics. According to rational theoretical approaches, successful comprehension requires that comprehenders closely align their knowledge with these statistics.

Evidence for adaptation in predictive sentence processing is mixed. For example, Brothers et al. (2017) observed greater facilitation for predictable words (e.g., "spider" following "The web had been spun by the large…" vs. "Alex said he wanted to watch the large…", which they captured through reading times) when participants were exposed to higher vs. lower proportions of predictable filler sentences. Moreover, no facilitation was observed when no filler sentences were predictable. These findings suggest that comprehenders may adapt their predictions based on experience to facilitate processing. According to rational theoretical approaches, disengaging predictive sentence processing may reflect a rational response when the reliability of comprehenders' beliefs is low and prediction error is high. Similarly, Zhang et al. (2023) replaced predictable words with semantically similar (i.e., vs. dissimilar) words and observed an attenuated event-related potential (ERP) response when participants were exposed to predictable but not incongruous filler sentences. However, at odds with both findings, Zhang et al. (2019) observed a similar ERP response on predictable (i.e., vs. unpredictable) words when participants were exposed to either predictable or incongruous filler sentences. These latter findings suggest that predictive sentence processing may be surprisingly insensitive to this experience.

An important limitation of many studies of adaptation (e.g., Brothers et al., 2017; Dave et al., 2021; Zhang et al., 2019, 2023) is that predictive behaviours were not captured directly. For example, Brothers et al. (2017) assessed reading times in response to predictable words (e.g., also see the ERP responses in Zhang et al. 2019, 2023), capturing processes after (e.g., which may not reflect prediction) and not before participants encountered this predictable input. Rather, particularly compelling evidence for predictive sentence processing is provided by studies measuring participants' behaviours before they encounter the predictable input. For example, the visual world paradigm (e.g., Tanenhaus et al., 1995) has been widely used to capture predictive eye movements. Altmann and Kamide's (1999) participants heard sentences like "The boy will eat the..." while viewing visual arrays with objects like a cake and other inedible distractors. Altmann and Kamide (1999) observed predictive eye movements to the cake following "eat" (i.e., vs. "move…"), suggesting that participants generated semantic predictions about the direct object based on the verb's selectional restrictions (e.g., edible). Crucially, Li et al. (2019) observed predictive eye movements (e.g., to a library when hearing "To borrow books, he followed the path to…") when participants were exposed to predictable but not semantically atypical filler sentences. Like Brothers et al. (2017) and Zhang et al. (2023), and consistent with rational theoretical approaches, these findings support the adaptivity of predictive sentence processing.

Alongside the visual world paradigm, ERP has also been used to capture predictive behaviours. For example, Van Wonderen and Nieuwland (2023) measured ERP responses on gender-marked articles before predictable nouns, building on related findings from the literature (e.g., Wicha et al., 2004). Their participants read sentences in Dutch like, "De politieagenten hadden de verdachte opgepakt. Hij moest direct mee naar het bureau…" ("The police officers had arrested the suspect. He immediately had to come to the station…"). Based on this discourse and the consistent grammatical gender of the article "het" and noun "bureau" ("station"; e.g., vs. sentences replacing the neuter gender "het" with the common gender "de"), Van Wonderen and Nieuwland (2023) observed an attenuated ERP response on the "het" before "bureau". Crucially, they observed a similar (i.e., pre-nominal prediction) effect when participants were exposed to either predictable or unexpected filler sentences. At odds with Li et al. (2019) and rational theoretical approaches, these findings suggest that comprehenders' predictions may not be sensitive to this experience.

Relatedly, the literature distinguishes different varieties of prediction. Luke and Christianson (2016) draw a distinction between lexical prediction and graded prediction. Lexical prediction refers to the prediction of specific words and graded prediction refers to the prediction of linguistic features. Luke and Christianson (2016) found that predictable words were rare in text passages taken from a variety of sources, in contrast to the high cloze probability sentences that are typical in studies of prediction. Thus, comprehenders' ability to adapt their lexical predictions may be poorly developed because this is rare, which may account for evidence showing that comprehenders do not adapt. Prediction-by-association (e.g., Pickering & Gambi, 2018; also see Huettig, 2015; Kukona et al., 2011) links prediction to spreading activation among associated representations within memory. For example, Kukona et al.'s (2011) participants heard sentences like "Toby arrests..." while viewing visual arrays with Toby and objects like a policeman, crook and other distractors.

Kukona et al. (2011) observed predictive eye movements during "arrest" to both the crook, which was related to the verb and a predictable direct object, and policeman, which was related to the verb but an unpredictable direct object. Thus, participants' predictions were not extinguished by conflicting information (e.g., see also Kamide & Kukona, 2018). Similarly, comprehenders' predictions may not be extinguished by exposure to unexpected sentences (e.g., which conflict with their predictions), which may also account for evidence showing that comprehenders do not adapt.

The aim of the current research was twofold: first, to assess (i.e., conceptually replicate) comprehenders' ability to adapt their predictions using mouse cursor tracking; and second, to assess comprehenders' ability to generate adapted predictions (e.g., in contrast to typical predictions based on lexical prediction, etc.). To summarise, while adaptation is the focus of a growing empirical and theoretical literature, support for adaptation in predictive sentence processing is mixed. Li et al. (2019) provide perhaps the most compelling support, but Van Wonderen and Nieuwland's (2023) impressive sample (*N* = 200), alongside related controversy surrounding syntactic adaptation (e.g., Dempsey et al., 2020, 2024; Harrington Stack et al., 2018; Prasad & Linzen, 2021), suggests that (e.g., conceptual) replication is essential. In addition, comprehenders' ability to generate adapted predictions (i.e., vs. simply disengaging prediction) remains an unassessed aspect of adaptation. For example, while Li et al. (2019) observed reduced eye movements to predictable objects when participants were exposed to semantically atypical filler sentences, an advantage for typically dispreferred (i.e., unpredictable) objects (e.g., restaurant when hearing "To borrow books, he followed the path to…") was not observed. Thus, perhaps at odds with rational theoretical approaches, participants did not (e.g., rationally) generate adapted predictions (i.e., centred on typically dispreferred objects). Instead, participants simply disengaged predictive sentence processing.

The current research captured predictive sentence processing using mouse cursor tracking (e.g., Spivey et al., 2005; for applications in sentence processing, also see Farmer, Anderson et al., 2007; Farmer, Cargill et al., 2007; Kukona et al., 2022). For example, Kukona's (2023) participants heard sentences like "What the man will ride, which is shown on this page, is the..." while viewing visual arrays with objects like a bike and kite. Consistent with Altmann and Kamide (1999), Kukona (2023) observed predictive mouse cursor movements to the bike following “ride” (i.e., vs. “spot…”), suggesting that participants generated semantic predictions about the direct object noun based on the verb’s selectional restrictions (e.g., ridable). Building on Kukona (2023), the current research assessed participants’ ability to adapt these predictive (i.e., mouse cursor) behaviours in response to (i.e., verb-based semantic) cue reliabilities. In Experiment 1, to assess the rational disengagement of prediction, participants heard equal proportions of non-predictive (e.g., “spot … the bike.”), predictive (e.g., “ride … the bike.”) and anti-predictive (e.g., “fly … the bike.”) sentences while viewing visual arrays like Figure 1 (e.g., with objects like a bike and kite). In Experiment 2, to assess the rational generation of adapted predictions, participants heard equal proportions of non-predictive and anti-predictive sentences, and no predictive sentences. Anti-predictive sentences were incongruous/atypical, such that their direct object noun referred to the verb-unrelated object, which was inconsistent with the verb’s selectional restrictions. The research question under focus was whether the reliability of verb-based semantic cues, which was manipulated through the relative proportions of non-predictive, predictive and anti-predictive sentences in these experiments, affected participants’ predictive behaviours (e.g., in contrast to typical predictions based on verb selectional restrictions; Kukona, 2023).

If predictive sentence processing is rationally adaptive, participants in Experiment 1 were not expected to make predictive mouse cursor movements because selectional restrictions provided an unreliable cue that was as likely to be disconfirmed as confirmed (e.g., bike was as likely following the anti-predictive verb “fly…” as predictive verb “ride…”), while participants in Experiment 2 were expected to make predictive mouse cursor movements to verb-unrelated objects (e.g., bike following “fly…”) because selectional restrictions provided a reliably disconfirmed cue (e.g., bike was more likely than kite following the anti-predictive verb “fly…”). Finally, motivated by Dave et al. (2021), participants’ cognitive control was also measured. Dave et al. (2021) observed greater adaptation (i.e., to individual talkers who produced more vs. less predictable sentences, such as “oven” vs. “house” following “Jack forgot about the cookies baking in the…”) among participants with higher cognitive control, which was measured through the classic Stroop (1935) task (e.g., for related findings, also see Jongman et al., 2023; Nozari et al., 2016). If adaptation is linked to cognitive control, greater adaptivity was expected among participants demonstrating better Stroop performance.

Experiment 1

Experiment 1 assessed adaptation when verb-based semantic cues were unreliable and as likely to be disconfirmed as confirmed. Mouse cursor movements to targets (e.g., bike) vs. distractors (e.g., kite) were measured while participants heard equal proportions of non-predictive (e.g., "spot…"), predictive (e.g., "ride…") and anti-predictive (e.g., "fly…") sentences.

Method

Transparency and openness

The power calculation used to determine the sample size, all data exclusions, all manipulations and all measures in the study are reported. All data, analysis code and research materials are available at OSF (Kukona & Hasshim, 2023). This study's design and its analysis were not pre-registered.

Participants

Eight-five participants (age $M = 39.78$, $SD = 10.75$, 12 unreported; 49 female, 36 male) were recruited through Prolific (https://prolific.co). Participants were UK native English speakers with normal or corrected-to-normal vision and hearing. The sample enabled detection of a medium correlational Cohen's effect size (i.e., between mouse cursor trajectories and cognitive control; $r = .30$, power = .80, alpha = .05). This study received research ethics committee approval from the Faculty Research Ethics Committee, De Montfort University (ref: 3661).

Design and materials

Verb type (non-predictive, predictive and anti-predictive) was manipulated within participants. Participants were presented 36 visual arrays from Kukona (2023). A target (e.g., bike) and distractor (e.g., kite) object from Duñabeitia et al. (2018) was depicted in each visual array. The visual array used normalised coordinates ranging from -1 to 1 (e.g., the coordinates at the extreme top-left of the visual array were [-1, 1]), with objects sized 0.3 x 0.6 and centred at (±0.85, 0.70). Each visual array was linked to three verbs, whose selectional restrictions were satisfied by: (1) both the target and distractor objects, reflecting a non-predictive verb (e.g., "spot"); (2) the target but not distractor object, reflecting a predictive verb (e.g., "ride"); and (3) the distractor but not target object, reflecting an anti-predictive verb (e.g., "fly"). Latent semantic analysis (e.g., Landauer & Dumais, 1997) revealed that: (1) non-predictive verbs did not differ in their relatedness to target ($M = 0.12$, $SD = 0.08$) and distractor ($M = 0.14$, $SD = 0.08$) objects, $t(35) = -1.27$, $p = .21$; (2) predictive verbs were more related to target ($M = 0.44$, $SD = 0.17$) than distractor ($M = 0.08$, $SD = 0.08$) objects, $t(35) = 11.74$, $p < .001$; and (3) anti-predictive verbs were less related to target ($M = 0.07$, $SD = 0.08$) than distractor ($M = 0.43$, $SD = 0.16$) objects, $t(35) = -12.49$, $p < .001$.

A female native speaker of British English recorded only plausible sentences, which minimised speaker cues to implausibility and were cross-spliced to create the experimental stimuli. The plausible sentences recorded included targets as direct objects of non-predictive (e.g., "What the man will spot, which is shown on this page, is the bike") and predictive (e.g., "What the man will ride, which is shown on this page, is the bike") verbs, as well as distractors as direct objects of anti-predictive verbs (e.g., "What the man will fly, which is shown on this page, is the kite"). The latter two recordings (i.e., which both include predictable direct objects) were not used directly as experimental stimuli. Rather, the first recording was used in the non-predictive condition and the latter two recordings were cross-spliced with the first recording post-verb for use in the predictive and anti-predictive conditions, respectively (e.g., the recording of "which is shown on this page, is the bike" was cross-spliced from the non-predictive condition into both the predictive [after "ride"] and anti-predictive [after "fly"] conditions). Thus, the target was the direct object across all conditions, and it was a plausible direct object (i.e., of the verb) alongside the distractor in the non-predictive condition, it was the more plausible direct object in the predictive condition and it was the less plausible direct object in the anti-predictive condition. The full list of items is reported in the Appendix. Participants were presented one of three counterbalanced lists. Each visual array was presented once on each list and in all three conditions across lists. Each list included 36 visual arrays with one third presented in each condition, and no practice or filler trials.

Procedure

The experiment was created in PsychoPy (Peirce et al., 2019) and run on Pavlovia (https://www.pavlovia.org). The mouse cursor tracking procedure closely followed Kukona (2023). Participants were instructed to use a computer mouse to click on an icon centred at the bottom of the visual display (0, -0.85) to begin each trial, then they were provided a preview of the visual array for 500 ms before hearing the sentence and finally they were instructed to click on the (i.e., target) object referred to in each sentence. No other demands were placed on participants' responses (e.g., hand movements or mouse cursor trajectories). Participants were informed that sentences may describe odd scenarios. Trial order and object location were randomised.

In the subsequent Stroop task, participants were presented congruent words spelling out the font colour they were presented in (i.e., "blue" in blue font, "green" in green or "yellow" in yellow) and incongruent words spelling out different colours (i.e., "orange" in blue, "purple" in green or "red" in yellow). Words were uppercase Open Sans font height 0.20 centred at (0, 0) against a white background. Participants were instructed to use their keyboard on each trial to respond to words in blue font with 1, green with 2 and yellow with 3. Participants viewed a black fixation cross for one second before each word. Participants were presented 150 trials following 20 practice trials with feedback. Condition and font colour were randomised.

Results and discussion

Two participants with mouse cursor response accuracies below 90%, as well as two participants who used a touchscreen (i.e., as reflected in concentrated starting/ending coordinates), were excluded from the analyses. One trial (0.03%) in which a response was made before target word (e.g., "bike") onset was also excluded. Accuracies were high across non-predictive (M = 99.79%, SD = 1.30), predictive (M = 99.90%, SD = 0.93) and anti-predictive (M = 99.38%, SD = 2.20) conditions. Inaccurate trials and trials with log RTs more than 2.5 standard deviations above the global mean (2.17%) were also excluded from the mouse cursor trajectory analyses.

The trajectory analyses focused on trials in which participants were expected to have at least some prior exposure to the verb type manipulation by excluding trials from early in the experiment (i.e., the first quartile, reflecting trials 1-9). These early trials, which corresponded to participants' first exposures to the verb type manipulation, are referred to as exposure trials. Figure 2A depicts mean trajectories across the visual array by verb type. Trajectories were aggregated by dividing each trial into 101 normalised time slices and inverting the horizontal axis for target objects presented on the left (e.g., see Spivey et al., 2005), such that horizontal x coordinates toward 1 were toward the target object and -1 the distractor object. Motor movements of the hand are complex, and a complete description of participants' mouse cursor trajectories is outside the scope of the current analysis. Rather, motivated by the typical analytical approach applied to eye movements in the visual world paradigm (e.g., Altmann & Kamide, 1999), the current analysis focused on the difference in attraction to objects between baseline (e.g., non-predictive verbs) and experimental (e.g., predictive and anti-predictive verbs) conditions during a (i.e., predictive) time window that preceded the occurrence of target words in sentences and spanned hundreds to thousands of milliseconds. While this (i.e., relative, between-condition comparison of) attraction is typically based on fixation proportions in the visual world paradigm, in the current research this attraction was captured spatially using (i.e., horizontal) x coordinates. Relative deflections in x coordinates towards objects in experimental vs. baseline conditions were interpreted as reflecting the prediction of those objects in the former vs. latter. While mouse cursor trajectories reflect a rich data source and provide a range of spatiotemporal measures, the current research focused on x coordinates because this dimension simply and clearly distinguished targets and distractors spatially. Figure 2B depicts mean x coordinates by verb type in 250 ms time slices from mean verb onset. These plots show that participants stabilised (i.e., excluding exposure trials) on visually similar trajectories across the non-predictive, predictive and anti-predictive conditions.

Trajectories were compared by analysing mean x coordinates (i.e., horizontal) during the 2 second temporal window preceding target word onset, which captures predictive mouse cursor movements prior to hearing the direct object noun (i.e., approximately reflecting, "which is shown on this page"). Predictive x coordinates excluding exposure trials were submitted to a mixed effects model with a dummy coded fixed effect of verb type with non-predictive as baseline and random intercepts by items. Models were run throughout in R using lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017). Maximal models were simplified throughout when there were issues with fit. The analysis revealed a non-significant difference between the non-predictive (Predictive x coordinates M = -0.01, SD = 0.09) and both predictive (M = 0.01, SD = 0.11), *Est.* = 1.40×10^{-2}, SE = 1.40×10^{-2}, t(2108) = 1.00, p = .32, and anti-predictive (M = 0.00, SD = 0.12), *Est.* = 0.12×10^{-2}, SE = 1.38×10^{-2}, t(2107) = 0.09, p = .93, conditions. Predictive x coordinates on exposure trials were also submitted to a similar model with random intercepts by participants and items. The analysis revealed a significant difference between the non-predictive (M = -0.01, SD = 0.18) and both predictive (M = 0.06, SD = 0.22), *Est.* = 5.87, SE = 2.96, $t(697) = 1.98$, $p < .05$, and anti-predictive (M = -0.11, SD = 0.23), *Est.* = -9.57, SE = 3.15, $t(694) = -3.04$, $p < .01$, conditions. Thus, while trajectories were drawn toward target objects in the predictive condition and distractor objects in the anti-predictive condition before the target noun early in the experiment, they were not differentially drawn toward these objects with continuing exposure to the verb type manipulation.

In addition, trajectories were also compared across (all) trials by analysing trial as a predictor, which did not require a cut-off for exposure trials. Predictive x coordinates including all trials were submitted to another mixed effects model with a fixed effect of verb type. The model also included linear and quadratic effects of trial, such that the first trial was coded as zero (i.e., reflecting the intercept; trial - 1), alongside their interactions with verb type. Figure 3A depicts mean predictive x coordinates by verb type across trials, as well as model fits. Table 1 reports model results. Predictive and anti-predictive trials differed significantly from non-predictive trials, such that while participants made predictive mouse cursor movements to verb-related objects in early trials, trials later converged.

Finally, individual differences in trajectories were also analysed. Stroop accuracy was high in both the congruent (M = 98.21, SD = 2.46) and incongruent (M = 98.38, SD = 3.31) conditions. Inaccurate Stroop trials and Stroop trials with RTs below 200 ms or above 2500 ms were excluded from the analyses. To test for a Stroop effect, mean log RTs were submitted to a by-participants mixed effect model with a dummy coded fixed effect of congruency type with congruent as baseline and random intercepts by participants. The analysis revealed a significant difference between the congruent (M = 689.11, SD = 185.40) and incongruent (M = 732.76, SD = 197.22) Stroop condition, *Est.* = 0.06, SE = 0.01, $t(80) = 9.80$, $p < .001$, reflecting a typical Stroop effect. Stroop performance was captured as the mean RT difference between the incongruent and congruent condition, which was also centred. A normalised approach, which captured Stroop performance by dividing this difference by the mean RT in the congruent condition, revealed similar results and thus is not reported. To test for a link between adaptation and cognitive control, predictive x coordinates excluding exposure trials were submitted to another mixed effects model with fixed effects of verb type and Stroop performance, as well as their interaction, and random intercepts by items. The current approach focused on asymptotic behaviours following exposure trials (e.g., vs. analysing trial as a predictor), particularly given concerns related to model complexity and fit. The analysis revealed a non-significant effect of Stroop performance in the non-predictive condition, *Est.* = 0.96×10^{-2}, *SE* = 0.96×10^{-2}, $t(2124) = 1.00$, $p = .32$, as well as non-significant interactions in the predictive, *Est.* = -2.19×10^{-2}, *SE* = 1.37×10^{-2}, $t(2127) = -1.60$, $p = .11$, and anti-predictive, *Est.* = -0.96×10^{-2}, *SE* = 1.39×10^{-2}, $t(2126) = -0.69$, $p = .49$, conditions (these effects were also non-significant on exposure trials; all *p*s > .10). Likewise, correlations between Stroop performance and predictive x coordinates in the non-predictive ($r = .11$, $p = .33$), predictive ($r = -.14$, $p = .22$) and anti-predictive ($r = -.03$, $p = .80$) conditions were also non-significant.

In Experiment 1, while participants initially made predictive mouse cursor movements to verb-related objects (e.g., targets like bike following predictive verbs like "ride…" and distractors like kite following anti-predictive verbs like "fly…") in early (i.e., exposure) trials, their subsequent mouse cursor movements were indistinguishable across conditions following these early trials. Consistent with Kukona (2023; also see Altmann & Kamide, 1999), these (i.e., former) results reveal that participants can use verb selectional restrictions to generate semantic predictions. Consistent with Li et al. (2019), these (i.e., latter) results also reveal that participants can disengage prediction when their predictions are unreliable and as likely to be disconfirmed (e.g., "fly … the bike") as confirmed (e.g., "ride … the bike"). Novelly, these results also extend prior research by revealing that mouse cursor tracking is sensitive to adaptation. These results support rational theoretical approaches, suggesting that comprehenders rationally disengaged predictive sentence processing when the reliability of their beliefs was low and prediction error was high. However, in contrast to Dave et al. (2021), a link between adaptation and cognitive control was not observed.

However, disengaging prediction may not be the only rational response to changing language statistics. Rather, in contrast to unreliable cues that are as likely to be disconfirmed as confirmed, reliable disconfirmations may alternatively provide a rational basis for generating adapted predictions. To the contrary, while Li et al. (2019) and Zhang et al. (2019, 2023) presented atypical or incongruous sentences with at least some degree of reliability, their participants were not observed to generate adapted predictions (i.e., centred on atypical or incongruous outcomes). Experiment 2 modified Experient 1 to assess comprehenders' ability to adapt when cues are more reliable.

Experiment 2

Experiment 2 assessed adaptation when verb-based semantic cues were reliably disconfirmed. Mouse cursor movements to targets (e.g., bike) vs. distractors (e.g., kite) were measured while participants heard equal proportions of non-predictive (e.g., "spot…") and anti-predictive (e.g., "fly…") sentences, and no predictive sentences.

Method

Participants

Eight-five participants (age *M* = 37.53, *SD* = 10.28, 11 unreported; 55 female, 30 male) were recruited following the same criteria as Experiment 1.

Design, materials and procedure

Verb type (non-predictive and anti-predictive) was manipulated within participants. Participants were presented the same visual arrays and sentences as Experiment 1 except that they were not presented predictive sentences. Participants were presented one of two counterbalanced lists. Each visual array was presented once on each list and in both conditions across lists. Each list included 36 visual arrays with one half presented in each condition and no practice or filler trials. The procedure was otherwise identical to Experiment 1.

Results and discussion

Fifteen trials (0.49%) in which a mouse cursor response was made before target word onset, were excluded from the analyses. Mouse cursor response accuracies were high across non-predictive (*M* = 99.93%, *SD* = 0.60) and anti-predictive (*M* = 100.00%, *SD* = 0.00) conditions. Inaccurate trials and trials with log RTs more than 2.5 standard deviations above the global mean (1.51%) were also excluded from the mouse cursor trajectory analyses.

The trajectory analyses again focused on trials in which participants were expected to have at least some prior exposure to the verb type manipulation (i.e., excluding trials 1-9). Figure 4A depicts mean trajectories across the visual array and Figure 4B depicts mean horizontal x coordinates from mean verb onset. These plots show visually diverging trajectories in the non-predictive and anti-predictive conditions. Predictive x coordinates (i.e., prior to hearing the direct object noun) excluding exposure trials were submitted to a mixed effects model with a fixed effect of verb type and random intercepts and slopes by participants and items. The analysis revealed a significant difference between the non-predictive (*M* = 0.02, *SD* = 0.11) and anti-predictive (*M* = 0.28, *SD* = 0.28) conditions, *Est.* = 25.52 x 10^{-2}, *SE* = 3.35 x 10^{-2}, $t(90) = 7.63$, $p < .001$. Predictive x coordinates on exposure trials were also submitted to a similar model with random intercepts by participants and items. The analysis revealed a significant difference between the non-predictive (*M* = -0.01, *SD* = 0.11) and anti-predictive (*M* = 0.06, *SD* = 0.25) conditions, *Est.* = 10.10 x 10^{-2}, *SE* = 2.40 x 10^{-2}, $t(737) = 4.20$, $p < .001$. Thus, trajectories were drawn toward target objects in the anti-predictive condition across the experiment. In addition, trajectories were again compared across (all) trials by analysing trial as a predictor. Predictive x coordinates including all trials were submitted to a mixed effects model with fixed effects of verb type, linear and quadratic effects of trial and their interactions. Figure 3B depicts mean predictive x coordinates by verb type across trials, as well as model fits. Table 1 reports model results. Anti-predictive trials differed significantly from non-predictive trials, such that the former diverged from the later across trials.

Finally, individual differences in trajectories were also analysed. Stroop accuracy was high across the congruent (*M* = 98.39, *SD* = 2.38) and incongruent (*M* = 98.57, *SD* = 2.73) conditions. The analysis of RTs revealed a significant difference between the congruent (*M* = 674.84, *SD* = 161.56) and incongruent (*M* = 720.29, *SD* = 174.42) Stroop condition, *Est.* = 0.06, *SE* = 0.00, $t(84) = 13.33$, $p < .001$, reflecting a typical Stroop effect. To test for individual differences, predictive x coordinates excluding exposure trials were submitted to another mixed effects model with fixed effects of verb type and Stroop performance, as well as their interaction, and random intercepts and slopes by participants and items. The analysis revealed a non-significant effect of Stroop performance, *Est.* = -1.78 x 10^{-2}, *SE* = 1.13 x 10^{-2}, $t(82) = -1.58$, $p = .12$, and non-significant interaction with condition, *Est.* = -2.14 x 10^{-2}, *SE* = 3.12 x 10^{-2}, $t(83) = -0.69$, $p = .49$ (these effects were also non-significant on exposure trials; all *t*s < 1). Likewise, correlations between Stroop performance and predictive x coordinates in the non-predictive ($r = -.18$, $p = .10$) and anti-predictive ($r = -.14$, $p = .20$) conditions were also non-significant.

In Experiment 2, participants made predictive mouse cursor movements to verb-unrelated objects (e.g., targets like bike following anti-predictive verbs like “fly…”). Novelly, these results extend Experiment 1 and Li et al. (2019) by revealing that participants can generate adapted predictions when relevant cues are reliably disconfirmed. These results support rational theoretical approaches, such that comprehenders rationally adapted their predictions in response to changing but reliable language statistics. However, consistent with Experiment 1, a link between adaptation and cognitive control was not observed.

General discussion

Two mouse cursor tracking experiments assessed comprehenders’ ability to rationally adapt their predictions during sentence processing. In Experiment 1, participants hearing equal proportions of non-predictive (e.g., “spot … the bike”), predictive (e.g., “ride … the bike”) and anti-predictive (e.g., “fly … the bike”) sentences did not make predictive mouse cursor movements to verb-related objects (e.g., ride-bike or fly-kite), suggesting that they disengaged prediction because constraining verb-based semantic cues (i.e., selectional restrictions) were unreliable and as likely to be disconfirmed as confirmed. In Experiment 2, participants hearing equal proportions of non-predictive and anti-predictive sentences made predictive mouse cursor movements to verb-unrelated objects (e.g., fly-bike), suggesting that they generated adapted predictions because constraining verb-based semantic cues were reliably disconfirmed. In addition, links between mouse cursor movements and individual differences in cognitive control were not observed. These experiments yield two novel insights into predictive sentence processing: first, comprehenders adapted by either disengaging prediction or generating adapted predictions; and second, this adaptivity was captured by measuring motor movements of the hand.

Experiments 1 and 2 complement prior empirical and theoretical research. While Kukona (2023) observed predictive mouse cursor movements to verb-related objects (e.g., bike following “ride”) when constraining verb-based semantic cues were reliably confirmed (e.g., building on classic findings like Altmann & Kamide, 1999; also see the exposure trials in Experiment 1), Experiments 1 and 2 revealed differing mouse cursor patterns when the reliability of these cues was manipulated. These results mirror findings like Brothers et al. (2017), who observed differing reading time patterns, Zhang et al. (2023), who observed differing ERP patterns, and Li et al. (2019), who observed differing eye movement patterns, when cue reliabilities were likewise manipulated. Taken together, these findings provide compelling evidence of the flexible adaptivity of predictive sentence processing. Moreover, the current results are interpreted as supporting rational theoretical approaches. These approaches emphasise comprehenders' probabilistic knowledge, which is assumed to underpin predictions that maximise successful comprehension (e.g., see Kuperberg & Jaeger, 2016). However, when the reliability of comprehenders' beliefs is low and prediction error is high (i.e., potentially impeding rather than supporting comprehension), comprehenders are assumed to rationally adapt by disengaging prediction. Likewise, closely related adaptation was observed across Brothers et al. (2017), Zhang et al. (2023), Li et al. (2019) and Experiment 1, such that comprehenders consistently disengaged prediction, as reflected in the absence of facilitated reading times, attenuated ERP responses and predictive eye and mouse cursor movements.

Experiment 2 also provides novel evidence of comprehenders' ability to generate adapted predictions. Relatedly, Dempsey et al. (2020) emphasise the dual nature of syntactic adaptation, which has been observed to both facilitate and disrupt processing. For example, Fine et al. (2013) exposed participants to typically dispreferred reduced relative garden paths (e.g., "The experienced soldiers warned about the dangers conducted the midnight raid.") within a study session, and observed facilitated processing for these constructions alongside disrupted processing for their typically preferred main verb counterparts (e.g., "The experienced soldiers warned about the dangers before the midnight raid."), although the latter is not without controversy (e.g., see Dempsey et al., 2020, 2024; Harrington Stack et al., 2018). While not perfectly analogous, Experiment 1 and prior studies (e.g., Brothers et al., 2017; Zhang et al., 2023; Li et al., 2019) are reflective of the latter, such that prediction of typically preferred (i.e., predictable) representations was disrupted. In contrast, Experiment 2 is reflective of the former, such that prediction of typically dispreferred (i.e., unpredictable) representations was facilitated.

Helpfully, predictive coding provides a mechanistic framework for explaining the current results. This framework assumes that cognition depends on the interaction of bottom-up sensory signals and top-down predictions throughout the neural hierarchy. Lupyan and Clark (2015) also hypothesise that language provides a flexible tool for tuning both what top-down predictions are recruited as well as how these top-down predictions are weighted. This flexibility is generally compatible with linguistic adaptation and may also distinguish the results of Experiments 1 and 2. Based on the predictive coding framework, participants in Experiment 1, whose predictive mouse cursor movements were not attracted to typically preferred over dispreferred objects, may have (i.e., adaptively) re-weighted the bottom-up sensory signal over top-down predictions because prediction error was persistently high. In contrast, participants in Experiment 2, whose predictive mouse cursor movements were attracted to typically dispreferred over preferred objects, may have (i.e., adaptively) recruited alternative top-down predictions (e.g., reflecting verb unrelatedness) because this reduced prediction error. Thus, reducing prediction error may provide a rational basis for guiding behaviour, linking predictive coding to rational theoretical approaches. Relatedly, an important distinction between studies of syntactic adaptation and the current experiments is that participants in the former were exposed to relevant linguistic structures (e.g., relative clauses) repeatedly. In contrast, the current participants were only exposed to specific verbs (e.g., fly) once. Thus, the re-organisation of lexical-semantic knowledge (e.g., the strengthening of fly-bike and weakening of fly-kite) cannot explain the current results because participants did not re-engage with this (i.e., verb-specific) information. An alternative possibility is that participants may be unable to adapt their predictions, but they were able to strategically delay their behavioural responses (e.g., until the occurrence of target words in sentences). However, while this possibility explains the results of Experiment 1, it cannot explain the results of Experiment 2, in which participants' attraction to typically dispreferred objects was not delayed. A variant of this possibility is that participants may be unable to adapt their predictions, but they were able to strategically use these predictions (i.e., of typically preferred objects) in Experient 2 to modify their subsequent behavioural responses. However, Experiment 2 provided no evidence that within trials, participants were initially attracted to typically preferred objects (e.g., kite when hearing "fly"; see Figure 4B) and only subsequently attracted to typically dispreferred objects (e.g., shifting their bias from kite to bike within trials). Rather, compatible with the predictive coding framework, we conjecture that these results are explained by adaptations to the recruitment and weighting of top-down predictions.

Perhaps surprisingly, comprehenders have not previously been shown to generate adapted predictions. According to rational theoretical approaches, adapted predictions that centre on typically dispreferred (i.e., unpredictable) representations may reflect a rational response when comprehenders' beliefs are reliably disconfirmed. In fact, an important similarity between Experiment 2 and prior studies (e.g., Brothers et al., 2017; Zhang et al., 2023; Li et al., 2019), but not Experiment 1 (i.e., which was maximally unreliable), is that constraining cues were disconfirmed with at least some degree of reliability. However, Brothers et al. (2017) and Zhang et al. (2023) did not observe facilitated reading times or attenuated ERP responses for typically dispreferred words when participants were exposed to lower proportions of predictable filler sentences, which would mirror the predictive mouse cursor movements to typically dispreferred objects in Experiment 2. We conjecture that this difference may depend on the visual arrays in Experiment 2, which provided considerable constraint, such that the target (e.g., bike) was the only typically dispreferred object (e.g., following "fly"). In contrast, Brothers et al. (2017) and Zhang et al. (2023) provided considerably less constraint, such that many words would typically be dispreferred (e.g., following "The web had been spun by the large…"). Likewise, Li et al. (2019) did not observe predictive eye movements to typically dispreferred objects when participants were exposed to lower proportions of predictable filler sentences, which would again mirror Experiment 2. While Li et al. (2019) did present constraining visual arrays, we conjecture that this difference may depend on their graded reliabilities. In other words, while the constraining cues were always disconfirmed in Experiment 2 (i.e., with 100% reliability), they were only disconfirmed 75% of the time in Li et al. (2019), perhaps diverging towards Experiment 1. Given these factors, it may have been impractical for comprehenders to adapt their predictions in these studies, and thus simply disengaging prediction was the most rational response. Nevertheless, an important direction for future research will be to assess the extent to which adapted predictions require high constraint and/or reliability.

The current results also contrast with the inflexibility observed by Van Wonderen and Nieuwland (2023). Again, they observed a similar pre-nominal prediction effect when participants were exposed to either higher or lower proportions of predictable sentences, which suggests that their participants did not rationally adapt their predictions. The current experiments diverge from Van Wonderen and Nieuwland (2023) along a range of dimensions, and thus definitively accounting for this difference is perhaps outside the scope of the current research, but one contrast worth highlighting is between semantic and syntactic prediction. Not only are distinct mechanisms hypothesised to underpin predictive sentence processing (e.g., Huettig, 2015), but there is also growing evidence of important processing differences. For example, while semantic prediction (e.g., Altmann & Kamide, 1999) is supported by an especially compelling empirical literature, evidence of phonological prediction is limited (e.g., see Nieuwland et al., 2018). Interestingly, evidence for adaptation in predictive sentence processing comes from studies focused on semantic prediction, including Brothers et al. (2017; e.g., web-spider), Zhang et al. (2023; e.g., Valentine's-roses) and Li et al. (2019; e.g., books-library), as well as Experiments 1 and 2 (e.g., ride-bike). In contrast, Van Wonderen and Nieuwland (2023) focused at least in part on syntactic prediction, as reflected in the grammatical gender of their articles and nouns (e.g., the consistent grammatical gender of the neuter gender "het" and noun "bureau" vs. common gender "de"; but also see the semantics of "politieagenten" and "bureau" [police officers-station]). Thus, we conjecture that semantic prediction may be uniquely adaptive, perhaps reflective of the considerable constraint provided by semantics (e.g., vs. grammatical gender), and that this adaptivity may not apply across other forms of prediction. However, a crucial direction for future research will be to systematically compare adaptation across different forms of prediction. Relatedly, another contrast worth highlighting concerns plausibility. The unexpected fillers in Van Wonderen and Nieuwland (2023) are described as "somewhat plausible or at least not incoherent", which contrasts with the current anti-predictive sentences (e.g., also see Li et al., 2019; Zhang et al., 2019, 2023). Many of the current sentences were highly implausible (e.g., impossible, at least literally, like "milk … the guitar"). Potentially, implausibility may provide an especially salient cue for adaptation. Thus, another crucial direction for future research will be to assess adaptation in response to more vs. less plausible sentences.

Finally, as a secondary focus of the current research, neither Experiment 1 nor 2 revealed links between adaptation and cognitive control. On the one hand, these results contrast with Dave et al. (2021), raising questions about the role of cognitive control in adaptation. On the other hand, these results do not compellingly resolve this issue and rather invite further research. A growing literature links predictive sentence processing to a range of cognitive individual differences, such as vocabulary knowledge (e.g., Borovsky et al., 2012; Borovsky & Creel, 2014; Hintz et al., 2017; Kukona et al., 2016; Mani & Huettig, 2012; Peters et al., 2018; Rommers et al., 2015; Sommerfeld et al., 2023). An important parallel between the Stroop task and the current experiments is that participants were required to bias their attention away from irrelevant information, such as the meaning of a word like "blue" or a predictable object like bike following "ride", and towards relevant information, such as the colour a word like "blue" is presented in or an unpredictable object like bike following "fly". Within the predictive coding framework, a potential hypothesis is that cognitive control may influence the recruitment and weighting of top-down predictions. However, one limitation of the current experiments is that the sample size was motivated generically by Cohen's effect sizes, which is not without its issues (e.g., Gignac & Szodorai, 2016). For example, Gignac and Szodorai (2016) reclassify a medium Cohen's effect size (i.e., r = .30) as a relatively large effect size based on the individual differences literature, and thus power reflects an important consideration for future research. Another limitation of the current frequentist approach is that it cannot provide support for a null effect, and thus a Bayesian approach may reflect a compelling alternative. Finally, James et al. (2018) highlight issues with capturing individual differences in sentence processing (e.g., reliability), and thus an experimental cross-task adaptation approach to cognitive control (e.g., Hsu & Novick, 2016) may reflect a compelling alternative.

In conclusion, the current experiments provide novel insight into the flexible adaptivity of predictive sentence processing. As predicted by rational theoretical approaches, participants were observed to adapt their predictions when constraining verb-based semantic cues were not reliably confirmed. However, alongside consistent (e.g., Brothers et al., 2017; Zhang et al., 2023; Li et al., 2019) and conflicting (e.g., Van Wonderen & Nieuwland, 202; Zhang et al., 2019) findings in the literature, we conjecture that this adaptivity may not apply to all forms of prediction. In addition, this adaptivity may depend on factors including the degree of reliability and plausibility. The current experiments also highlight the sensitivity of online mouse cursor tracking to the moment-by-moment processes underpinning language comprehension. Moreover, the current research suggests that online mouse cursor tracking provides a powerful tool for assessing adaptation beyond (e.g., the limitations of) the lab.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583-609.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology, 50*(5), 1600-1613.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417-436.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language, 93*, 203-216.

Dave, S., Brothers, T., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2021). Cognitive control mediates age-related changes in flexible anticipatory processing during listening comprehension. *Brain Research, 1768*, 147573.

Dempsey, J., Liu, Q., & Christianson, K. (2024). Syntactic adaptation leads to updated knowledge for local structural frequencies. *Quarterly Journal of Experimental Psychology, 77*(2), 363-382.

Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(10), 1906-1921.

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*(4), 570-595.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science, 31*(5), 889-909.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS ONE, 8*(10), e77661.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1-76.

Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition, 46*, 864-877.

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(9), 1352.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science, 27*(4), 572-582.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118-135.

James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions?. *Journal of Memory and Language, 102*, 155-181.

Jongman, S. R., Copeland, A., Xu, Y., Payne, B. R., & Federmeier, K. D. (2023). Older adults show intraindividual variation in the use of predictive processing. *Experimental Aging Research, 49*(5), 433-456.

Kaan, E., Futch, C., Fuertes, R. F., Mujcinovic, S., & de la Fuente, E. Á. (2019). Adaptation to syntactic structures in native and nonnative sentence comprehension. *Applied Psycholinguistics, 40*(1), 3-27.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2*(4), 647-670.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition, 124*(1), 66-71.

Kamide, Y., & Kukona, A. (2018). The influence of globally ungrammatical local syntactic constraints on real-time sentence comprehension: Evidence from the visual world paradigm and reading. *Cognitive Science, 42*(8), 2976-2998.

Kukona, A. (2023). Predictive sentence processing at speed: evidence from online mouse cursor tracking. *Cognitive Science, 47*(4), e13285.

Kukona, A., Braze, D., Johns, C. L., Mencl, W. E., Van Dyke, J. A., Magnuson, J. S., ... & Tabor, W. (2016). The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica, 171*, 72-84.

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119*(1), 23-42.

Kukona, A., Gaziano, O., Bisson, M. J., & Jordan, A. (2022). Vocabulary knowledge predicts individual differences in the integration of visual and linguistic constraints. *Language, Cognition and Neuroscience, 37*(6), 750-765.

Kukona, A., & Hasshim, N. (2023). Mouse cursor trajectories capture the flexible adaptivity of predictive sentence processing [Data set]. OSF. https://osf.io/ndsmf

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience, 31*(1), 32-59.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

Li, Q., Chen, X., Su, Q., Liu, S., & Huang, J. (2019). Adapt retrieval rules and inhibit already-existing world knowledge: Adjustment of world knowledge’s activation level in auditory sentence comprehension. *Language and Cognition, 11*(4), 645-668.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22-60.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science, 24*(4), 279-284.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843-847.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *Elife, 7*, e33468.

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review, 23*, 1942-1953.

Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(7), 1156-1172.

Peters, R. E., Grüter, T., & Borovsky, A. (2018). Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied Psycholinguistics, 39*(6), 1439-1469.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002-1044.

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics, 77*, 720-730.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393-10398.

Sommerfeld, L., Staudte, M., Mani, N., Kray, J. (2023). Even young children make multiple predictions in the complex visual world. *Journal of Experimental Child Psychology, 235*, 105690.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Van Wonderen, E., & Nieuwland, M. S. (2023). Lexical prediction does not rationally adapt to prediction error: ERP evidence from pre-nominal articles. *Journal of Memory and Language, 132*, 104435.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*(2), 250-271.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272-1288.

Zhang, W., Chow, W. Y., Liang, B., & Wang, S. (2019). Robust effects of predictability across experimental contexts: Evidence from event-related potentials. *Neuropsychologia, 134*, 107229.

Zhang, W., Dong, J., Duan, X., Zhang, Y., Gao, X., Zhen, A., ... & Yan, H. (2023). Prediction of semantic features is modulated by global prediction reliability: Evidence from the N400 effect. *Journal of Neurolinguistics, 65*, 101109.

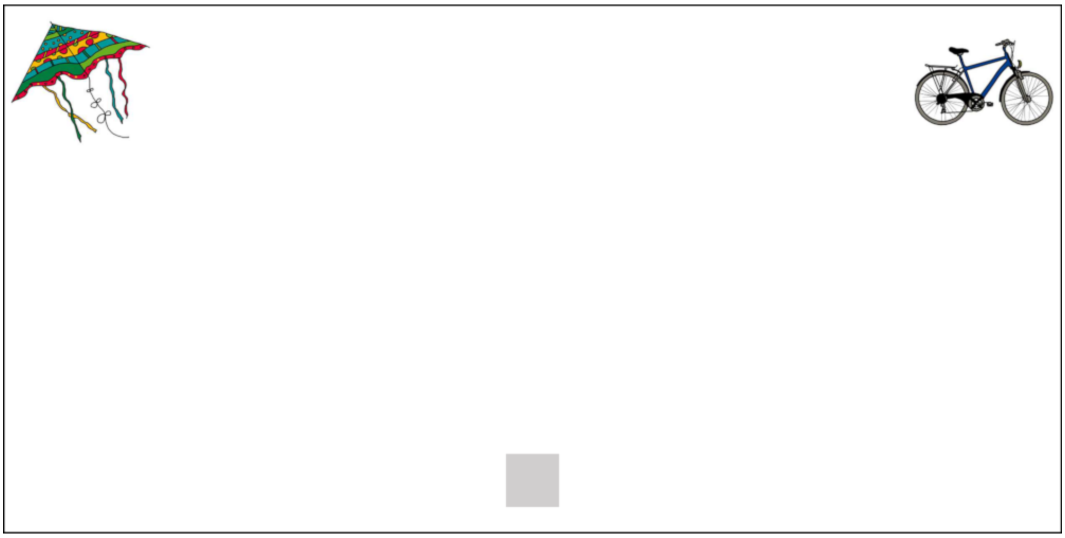
Table 1

Analysis of predictive horizontal mouse cursor movements (i.e., x coordinates) by verb type across (all) trials

Experiment 1			
Fixed effect	***Est.* (*SE*)**	***t***	***p***
Intercept	-0.69 (2.42)	-0.28	.78
Pred	13.28 (3.54)	3.75	< .001
Anti	-16.05 (3.74)	-4.29	< .001
Trial	0.06 (0.33)	0.19	.85
$Trial^2$	0.00 (0.01)	-0.23	.82
Pred x Trial	-1.64 (0.48)	-3.43	< .001
Anti x Trial	1.67 (0.49)	3.40	< .001
Pred x $Trial^2$	0.04 (0.01)	3.25	< .01
Anti x $Trial^2$	-0.04 (0.01)	-2.78	< .01
Experiment 2			
Fixed effect	***Est.* (*SE*)**	***t***	***p***
Intercept	-2.29 (2.36)	-0.97	.33
Anti	0.68 (4.11)	0.17	.87
Trial	0.38 (0.29)	1.29	.20
$Trial^2$	-0.01 (0.01)	-0.91	.37
Anti x Trial	2.17 (0.43)	5.07	< .001
Anti x $Trial^2$	-0.04 (0.01)	-3.46	< .001

Figure 1

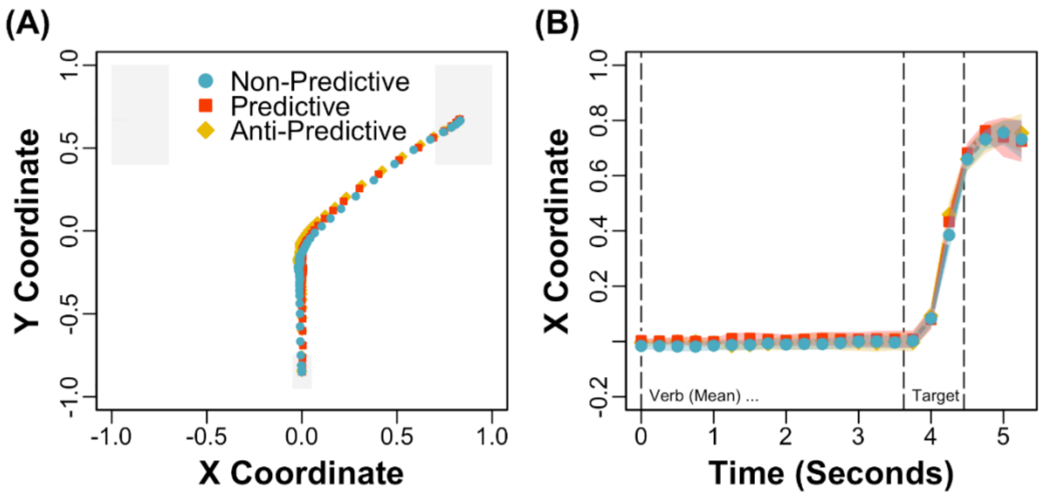
Example visual array depicting a target bike and distractor kite



Note. Participants heard "What the man will spot, which is shown on this page, is the bike" in the non-predictive condition, "What the man will ride, which is shown on this page, is the bike" in the predictive condition and "What the man will fly, which is shown on this page, is the bike" in the anti-predictive condition.

Figure 2

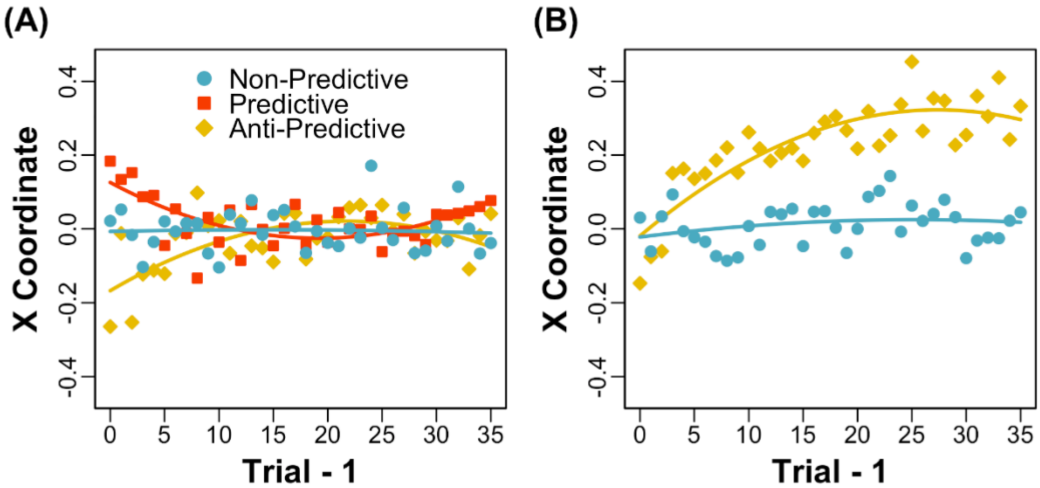
Time-normalised mean mouse cursor trajectories across the visual array (A) and mean (shaded bands show 95% CIs) horizontal mouse cursor movements (i.e., x coordinates) from mean verb onset (B) by verb type in Experiment 1 excluding exposure trials



Note. Participants heard equal proportions of non-predictive (e.g., “spot … the bike”), predictive (e.g., “ride … the bike”) and anti-predictive (e.g., “fly … the bike”) sentences.

Figure 3

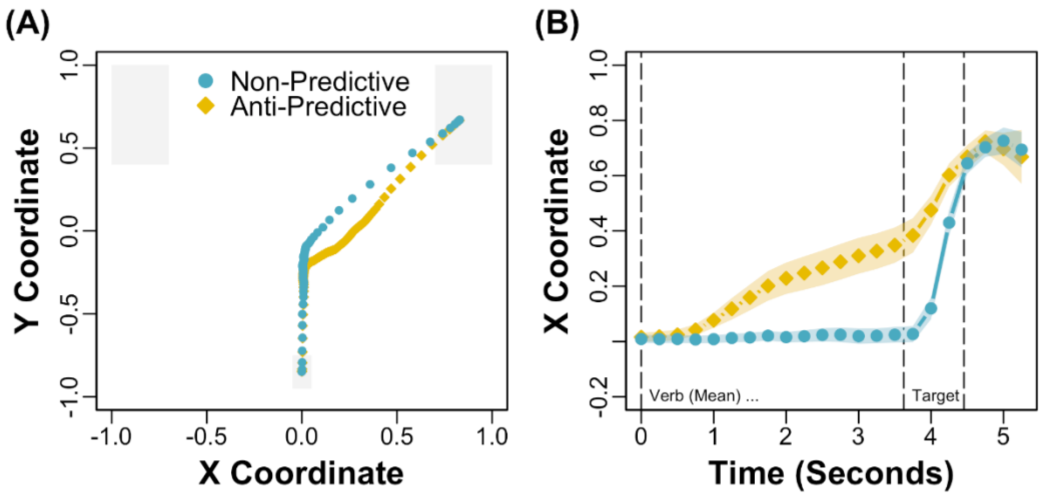
Mean predictive horizontal mouse cursor movements (i.e., x coordinates) by verb type across trials in Experiments 1 (A) and 2 (B)



Note. Lines depict model fits.

Figure 4

Time-normalised mean mouse cursor trajectories across the visual array (A) and mean (shaded bands show 95% CIs) horizontal mouse cursor movements (i.e., x coordinates) from mean verb onset (B) by verb type in Experiment 2 excluding exposure trials



Note. Participants heard equal proportions of non-predictive (e.g., "spot … the bike") and anti-predictive (e.g., "fly … the bike") sentences, and no predictive sentences.

Appendix

Table A1 reports the target and distractor objects from Experiments 1 and 2, alongside the corresponding non-predictive, predictive and anti-predictive verbs.

Table A1

Objects and verbs

Target	Distractor	Non-predictive	Predictive	Anti-Predictive
bell	shower	glance at	ring	rinse off in
medal	airport	look at	win	arrive into
roof	knife	notice	climb on	cut with
chocolate	helicopter	see	melt	pilot
drum	pencil	spot	play	draw with
pool	curtain	stare at	dive into	hang
pizza	keyboard	view	slice	type with
flag	net	watch	salute	fish with
battery	kitchen	ask about	recharge	cook in
car	tomato	chat about	drive	puree
potato	scissors	discuss	bake	sharpen
tree	bible	enquire about	prune	preach from
baby	doughnut	hear about	cradle	frost
banana	sock	learn about	peel	stitch
dolphin	basket	speak about	swim with	weave
egg	beard	talk about	poach	shave
balloon	queen	think about	inflate	crown
chair	belt	wonder about	sit on	buckle
rope	hair	glance at	tie	brush
ice cream	wave	look at	freeze	surf
mountain	bomb	notice	hike up	disarm
compass	sink	see	navigate with	wash in
bike	kite	spot	ride	fly
gun	bed	stare at	shoot	sleep in
guitar	cow	view	strum	milk
camera	door	watch	focus	slam
king	microphone	ask about	dethrone	talk into
lion	spoon	chat about	hunt	stir with
piano	candle	discuss	tune	light
fork	ruler	enquire about	eat with	measure with
cage	bottle	hear about	lock	drink from
fruit	gym	learn about	ripen	exercise in
book	dragon	speak about	publish	slay
road	scarf	talk about	pave	knit
carrot	dress	think about	roast	hem
helmet	clown	wonder about	wear	laugh at