Magnitude Estimation Reveals Temporal Binding at Super-Second Intervals

Gruffydd R. Humphreys and Marc J. Buehner
Cardiff University

Several recent studies (e.g., Haggard, Aschersleben, Gehrke, & Prinz, 2002; Haggard & Clark, 2003; Haggard, Clark, & Kalegeras, 2002) have demonstrated a “Temporal Binding” effect in which the interval between an intentional action and its consequent outcome is subjectively shorter compared to equivalent intervals that do not involve intentional action. The bulk of the literature has relied on the “Libet Clock” (Libet, Gleason, Wright, & Pearl, 1983; but see also Engbert & Wohlschläger, 2007; Engbert, Wohlschläger, Thomas, & Haggard, 2007; Engbert, Wohlschläger, & Haggard, 2008). Here we demonstrate that Temporal Binding is a robust finding that can also be reliably achieved with a Magnitude Estimation procedure, and that occurs over intervals far greater than those previously explored. Implications for the underlying mechanisms are discussed.

Keywords: interval estimation, temporal binding, time perception
binding occurred. In addition, when participants observe the experimenter press the button, Temporal Binding still appears, but not when the button depresses of its own accord (Wohlschlager, Haggard, Gesierich, & Prinz, 2003) or is depressed by a rubber hand (Wohlschlager, Engbert, & Haggard, 2003), because neither presumably displays intentionality. This dependence of temporal binding on intentionality was repeated when the voluntary action was perceived to be involuntary through hypnotic suggestion (Haggard, Cartledge, Dafydd, & Oakley, 2004). The enforcing of an involuntary action via Transcranial Magnetic Stimulation (Haggard, Clark, & Kalogeras, 2002; Haggard & Clark, 2003) produced a temporal repulsion: relative to baseline participant judgments of action and effect onset were shifted away from each other in time. Furthermore, the magnitude of the effect has been shown to influence the extent of binding, with more intense tones eliciting a larger shift relative to baseline (Engbert, 2005).

A more general appraisal of the temporal binding effect was offered by Eagleman and Holcombe (2002), who argued that the causal relationship between action and effect is the crucial ingredient. Harking back to Hume (1739/1888) they pointed out that temporal contiguity is a well-established cue towards causality. They conjectured that the mind utilizes the temporal contiguity cue in a bidirectional manner: Not only are we likely to consider contiguous event pairings as causal, but we may also consider causal event pairings as more contiguous. In situations in which there is some uncertainty in temporal measurement, the binding together of two events that are known to be causally related can be seen as an adaptive process that reconstructs the relationship between these events (Stetson, Xu, Montague, & Eagleman, 2006). Note that the causal analysis of temporal binding is not in contradiction with the intentional analysis. Rather, according to this view, intentional binding is simply a special case of causal binding, such that the cause happens to be an intentional action.

The finding of a TMS repulsion effect (Haggard, Clark et al., 2002) mentioned above might appear to contradict Eagleman and Holcombe’s (2002) analysis. However, close inspection of the experimental protocol in this condition reveals that in fact there was no causal relation between the TMS-induced muscle twitch and the beep that followed; rather, both the TMS stimulation and the delivery of the tone were controlled by an experimental program. Presumably, participants were acutely aware that no causal link existed between their muscle twitch and the tone that followed it, and hence no binding would be expected under either an intentional or a causal perspective.

This paper, however, is not designed to investigate the intentional/causality dichotomy. Our further use of the terms intentional and causal in this paper thus does not indicate a selective endorsement of either hypothesis. We intended to follow up on what we saw as a main shortcoming in the wealth of evidence in support of Temporal Binding outlined above: the almost exclusive usage of the Libet Clock paradigm. The method itself has over many years received much in the way of both support and criticism (Libet, 1985, 2002; Pickett & Miller, 2007). In the current context of temporal binding, we see two main limitations or problems. Firstly, the effect could simply be an artifact of the method. For example, the Libet Clock has previously been subject to some criticism based on the “Flash Lag” effect (Nijhawan, 1994), in which a moving stimulus and a flash appearing at the same location are judged to be spatially separate. It has been argued that a fundamental flaw of the Libet Clock may be its vulnerability to this effect (e.g., Gomes, 2002, Pockett, 2002). Although it may be that the appropriation of baselines acts to cancel out any flash-lag specific effects pertaining to the reality of Temporal Binding (cf. Moore & Haggard, 2008) this contention is sufficient to lead us to seek a new paradigm. Secondly, the Libet Clock examines the timing of one event per trial relative to a baseline, and does not measure the relative interval between events. Clearly, if one is interested in studying temporal binding, a more direct measure would be to study changes in the perceived length of elapsed time between events. The Libet Clock method at best offers a very indirect route to obtain this measure.

Therefore, our first goal was to investigate whether Temporal Binding is a robust phenomenon that can be demonstrated with methods other than the Libet Clock; more specifically, we employed a free numerical temporal estimation method to measure the subjective interval between intentional actions and their subsequent effects. A recent series of experiments have demonstrated a binding effect with a magnitude estimation method at intervals of 200 to 300 ms (Engbert & Wohlschlager, 2007; Engbert, Wohlschlager, & Haggard, 2008; Engbert, Wohlschlager, Thomas, & Haggard, 2007). However, we intended to replicate temporal binding with a magnitude estimation procedure, and the intervals used in Haggard, Aschersleben et al.’s (2002) original report (Experiment 1a). Haggard, Aschersleben et al. investigated binding at intervals up to 650 ms, and reported that the size of the binding effect diminished as the temporal interval increased. This could either reflect a natural limitation of the binding effect, or be rooted in artificial constraints of the Libet Clock method. Consequently, we also aimed to investigate whether binding occurs at intervals longer than 650 ms (Experiments 1b-e). To foreshadow our results: we were surprised to find that Experiment 1a suggested that the temporal binding effect increased with interval size. Based on Haggard, Aschersleben et al.’s results, we had predicted a decrease. Inspired by this surprise result, we proceeded to push Temporal Binding to its limits, and increased the temporal intervals under investigation up to 4 s, where we still found reliable binding.

Experiments 1a-e

All Experiments involved the same basic method. Participants were presented with Operant and Observational trials. On Operant trials, participants pressed a button at a time of their choosing, which resulted in the delivery of a 100-ms, 1-kHz pure tone after a set interval. On Observational trials, participants did not perform any intentional or causal actions; instead, they were presented with an audible click, which was followed by the same 100-ms, 1-kHz pure tone after a set interval. The click in these trials was recorded from the button box used on Operant trials, and thus was identical to the noise made by the microswitch when participants pressed the button in the Operant trials. At the end of each trial, participants had to estimate the length of the interval between their button press and the tone (Operant trials), or the click and the tone (Observational trials). Because Operant trials involved intentional causal action and Observational trials did not, we hypothesized that participants would judge the interevent intervals to be shorter in Operant trials than in relative Observational trials.
Method

Participants. Sixteen different Cardiff University undergraduate students participated in each of Experiments 1a-d, and seventeen in Experiment 1e, either to partially fulfill a course requirement, or to receive £4. Each participant served in one experiment only, so each experiment had a different cohort of participants.

Materials and apparatus. The experiments were conducted on an Apple iMac, and programmed with Psyscope (Cohen, Mac-whinney, Flatt, & Provost, 1993); the psycscope “Button Box” was used for timing of stimulus delivery and as an input device in the Operant trials. Temporal estimates were entered via the keyboard. Each experimental trial contained two events, separated by an interval. The first was a button press in Operant, and a click in Observational trials; the second was a 100-ms, 1-kHz pure tone. A green or red fixation cross at the centre of the screen denoted Operant or Observational trials, respectively. Crosses were removed at the beginning of the first event of each trial.

Design and procedure. The factors Interval (see below) and Trial Type (Observational vs. Operant) were factorially combined in a within-subjects procedure. The Intervals used across the four experiments were as follows. Experiment 1a: 150, 250, 350, 450, 550, and 650 ms; Experiment 1b: 750, 850, 950, 1,050, 1,150, and 1,250 ms; Experiment 1c: 0, 250, 500, 750, 1,000, 1,250, 1,500, 1,750, and 2,000 ms; Experiments 1d and 1e: 0, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, and 4,000 ms.

Operant and Observational trials were blocked so that participants experienced one trial of each interval per block. The order of intervals within a block was random, and Operant and Observational blocks alternated. Each participant worked on 10 Operant and 10 Observational blocks, thus providing 20 separate temporal estimates for each interval, 10 of which involved Operant, and 10 Observational trials. The nature of the first block (Operant vs. Observational) was counterbalanced between participants.

Each interval began with a display of a fixation cross at the center of the screen. The cross was green on Operant and red on Observational trials. On Operant trials, pressing the green button on the response box cleared the fixation cross from the screen and triggered the relevant interval, after which the tone was delivered. On Observational trials, the red cross remained on the screen for a random interval between 1,500 and 2,000 ms, after which the click was presented and the cross disappeared; the pure tone was then delivered after the appropriate interval. Although participants were asked to judge the interval between events, it was possible that participants would begin interval timing sometime during this (~120 ms) click stimulus. As such, we adopted a highly conservative timing criterion, and programmed the experiment such that the timing of this interval commenced at the beginning of the presented click, thus potentially shortening the Observational interval for our participants, and working against our hypothesis.

Each trial ended with an onscreen prompt to provide an estimate for the interevent interval. The response range was restricted from 0 to 999 ms in Experiment 1a, from 0 to 1,999 ms in Experiment 1b, from 0 to 2,000 ms in Experiment 1c, from 0 to 4,000 ms in Experiment 1d, and from 0 to 5,000 ms in Experiment 1e. Once the Return key was depressed, the response was entered the next trial began after a random interval between 750 to 1,000 ms.

Participants were informed that they would be partaking in a study of time perception. After giving written consent to participate in the study participants were provided with a general outline of the experimental procedure, followed by written instructions specific to Operant or Observational blocks. At the start of Operant blocks, participants were informed that the appearance of the tone was wholly dependent on their actions: Pressing the green button on the Button Box would produce a tone after a set interval. Instructions suggested that participants could delay their button press: The absence of a tone during this duration demonstrated that the delivery of the tone was wholly dependent on their action. Before Observational blocks, instructions emphasized that participants were not required to press the key but would passively observe two unrelated events. These instructions were presented at the beginning of each block, to facilitate clear discrimination between Operant and Observational blocks.

Results and Discussion

Median temporal judgments. Each participant returned 10 temporal judgments per Trial Type × Interval combination in each Experiment. Participants’ median temporal judgments served as the primary units of analysis, and are displayed in Figures 1 through 5. Median judgments more than two standard deviations from the mean of all median judgments for a particular Trial Type × Interval combination were considered outliers, and participants contributing one or more outliers were removed from the analyses. This criterion led to the exclusion of 2 participants in Experiment 1a, 1 participant in Experiment 1b, 3 participants in Experiment 1c, 3 participants in Experiment 1d, and 4 participants in Experiment 1e. All statistical analyses adopted a significance level of 0.05 except where otherwise noted.

Inspection of Figures 1 through 5 shows that participants clearly distinguished between the various intervals, and suggests that Operant intervals were consistently judged shorter than equivalent Observational intervals. Furthermore, it appears that this difference increased as a function of interval duration in Experiments 1a, b, c, and e. Statistical analyses corroborate this impression. We performed repeated measures Trial Type × Interval ANOVAs for each experiment separately and found consistent and robust effects of Trial Type, Experiment 1a: F(1, 13) = 10.60, MSE = 35022.11, η² = .45; Experiment 1b: F(1, 14) = 19.58, MSE = 49923.16, η² = .58; Experiment 1c: F(1, 12) = 13.28, MSE = 85354.86, η² = .53; Experiment 1d: F(1, 12) = 5.44, MSE = 52229.87, η² = .31, and Experiment 1e: F(1, 12) = 5.36, MSE = 113482.62, η² = .039. This result indicates that judgments derived from Operant intervals were overall judged to be shorter than judgments from Observational

1 Experiment 1e is a replication of Experiment 1d, but with a wider response range, aimed at ruling out a concern raised by an anonymous reviewer that limiting the response range to the maximum interval (4,000 ms) might have biased the results. Our replication shows that this was not the case.
intervals. As would be expected, the main effect of Interval was also significant in each experiment, indicating that participants successfully discriminated different interval lengths, Experiment 1a: $F(5, 65) = 23.52, \text{MSE} = 18485.60, \eta^2 = .64$; Experiment 1b: $F(5, 70) = 24.12, \text{MSE} = 41883.41, \eta^2 = .63$; Experiment 1c: $F(8, 96) = 133.67, \text{MSE} = 47421.25, \eta^2 = .92$; Experiment 1d: $F(8, 96) = 217.38, \text{MSE} = 174139.16, \eta^2 = .95$; Experiment 1e: $F(8, 96) = 187.037, \text{MSE} = 212799.528, \eta^2 = .94$.

The slope from Operant interval estimations was significantly shallower than the slope from Observational intervals. The slope from Operant interval estimations was significantly shallower, 2 a steeper one, and 1 tie. However, the same intervals produced reliably different slopes (one-tailed) in Experiment 1e, $Z = 1.70$, with 11 participants returning a shallower slope, 7 a steeper one, and 1 tie. When data from both experiments were nonsignificant. In these cases, we also calculated the analyses based on just the significant regression models. The results do not change.

### Slope analyses

To better understand how the Operant/Observational effect changes with increasing intervals we additionally conducted Slope Analyses, as previously employed in experiments involving numerical estimation of duration (Penton-Voak, Edwards, Percival, & Wearden, 1996; Wearden, Edwards, Fakhri, & Percival, 1998). To this end, we calculated for each participant an individual regression slope across his or her median temporal judgments of the Operant and Observational intervals within a given experiment. We then compared Operant and Observational slopes with Wilcoxon Signed Rank tests, because of the nonnormal distribution of slopes.

The slope from Operant interval estimations was significantly lower slope and 3 a steeper one. The slopes in Experiment 1d did not differ reliably, $Z = 1.02$, with 8 participants returning a shallower slope, 7 a steeper one, and 1 tie. However, the same intervals produced reliably different slopes (one-tailed) in Experiment 1e, $Z = 1.70$, with 11 participants returning a shallower slope, 4 a steeper one, and 1 tie. When data from both Experiment

---

2 One anonymous reviewer expressed concern that when comparing Experiments 1a and 1b to Experiments 1c-e, the size of the interval range and the absolute interval values over which that range spans are partially confounded, and that because of this, our experiments cannot shed light on a possible contribution of range effects on participants estimates (e.g. Poulton, 1979). Range effects may well have biased participants’ judgments differently across these two groups of experiments. Importantly, however, any bias induced by range effects would have affected Operant and Observational conditions equally because they employed exactly the same intervals and ranges. Furthermore, since these conditions were varied within participants, our central finding of a robust main effect of Trial Type could not be compromised by different range effects across these studies.

4 The slope analyses reported here include all regression slopes, irrespective of whether each individual participant’s model reached significance. In Experiment 1a, one model was nonsignificant; and in Experiment 1b, five models were nonsignificant. In these cases, we also calculated the analyses based on just the significant regression models. The results do not change.

5 One participant was excluded from slope analysis because he or she returned the same judgment for each interval.
1d and e is merged, the overall analysis also indicates reliably shallower slopes for Operant than Observational intervals, $Z = 1.90$, with 19 participants returning a shallower, 12 a steeper slope, and 2 ties.

### General Discussion

We set out to replicate Haggard, Aschersleben et al.’s (2002) results with a magnitude estimation method. Experiment 1a dem-
Demonstrated that operant intervals were reliably underestimated compared to equivalent observational intervals within the interval range used by Haggard, Aschersleben et al. (up to 650 ms). Importantly, this result was obtained using a direct measure of interval length–magnitude estimation. Thus, the temporal binding effect reported by Haggard, Aschersleben et al. is an empirically robust phenomenon, that occurs not only when indirect measures (temporal shifts calculated from single-event subjective time-stamps) are employed, but also when participants have to report their temporal percepts directly. The robust nature of the effect and, more specifically, its apparent increase with interval size was at variance with Haggard, Aschersleben et al.’s finding of a decrease. Thus, we set out to investigate whether temporal binding might also occur over substantially longer intervals, and have demonstrated, for the first time, that binding is steady across intervals up to 4 s.

How can we reconcile this latter aspect of our results with earlier work of Haggard, Aschersleben et al. (2002) and Haggard, Clark et al. (2002) who reported a steady decrease in binding as the interval size increased from 250 to 650 ms? A tempting suggestion would be that the answer lies in the different experimental methods. Whereas the Libet Clock method is dependent on taking the judged onset of an event and comparing it to a baseline, the Temporal Judgment method offers a direct examination of the subjective interval. Directly examining the subjective interval between cause and effect thus might be more sensitive to the binding effect at longer intervals.

A more satisfying, and theoretically interesting explanation arises when we consider the potential processes underlying temporal binding. If temporal judgments reflect the processes of an internal clock with a pacemaker (see for example Wearden, 2001 for an overview), our results, and in particular the reliable differences in slopes, would suggest that the pacemaker runs slower during periods where people anticipate the consequences of their intentional causal actions. Although a stable difference between Operant and Observational trials could be explained by a delay in the switch of the accumulator at the onset of Operant trials, or an anticipatory, early switching off at the end of Operant trials, the steady increase and differences in slopes are more suggestive of different pacemaker speeds. A slower pacemaker would mean that fewer pulses are accumulated, which in turn implies that the Operant/Observational difference increases with longer intervals. Indeed the general pattern of our results in Figures 1 through 5 closely resemble the results of other articles in which the disparity in numerical estimations of the duration of two stimuli have been explained in terms of a modulation in the speed of a pacemaker (e.g. Penton-Voak et al., 1996; Wearden, 2008; Wearden et al., 1998). The timing literature lists various determinants of pacemaker speed, including stimulus modality, interval content, arousal, etc. (for an overview see Wearden, 2001), and it may well be that causality and/or intentionality will need to be added into this canon.

An additional explanation for our results, and in particular our finding of temporal binding at Super-Second intervals, would be that at least part of the effect is postdictive in nature. Moore and Haggard (2008) have argued that intentional binding involves both pre- and postdictive components. Their task manipulated the probability with which an intentional action led to the expected result. Using the Libet Clock, Moore and Haggard found that intentional actions were judged later relative to baseline in both high- and
low-contingency conditions on “successful” trials where a key press was followed by a tone, thus displaying the familiar perceptual shift underlying temporal binding. In contrast, when considering “unsuccessful” trials, where a key press did not produce the tone, the shift was apparent only in the high-contingency condition. Taken together, these results suggest that a strong expectation of the effect (in successful, and—critically—also in unsuccessful trials of the high-contingency condition), led to a predictive perceptual shift, and that the observation of the effect (in successful trials of the low-contingency condition) led to a postdictive perceptual shift. Moore and Haggard referred to the predictive component as sensory based, and to the postdictive component as inference based, and suggested that both are integrated in a Bayesian manner. This distinction between, and integration of, sensory and inferential processes is pertinent to our results.

Whereas sensory-based shifts are very short lived and might not outlast the execution of an action (cf. Moore & Haggard, 2008), there is no a priori reason why inferential shifts could not operate over longer timescales. The Libet Clock method, of course, taps largely into sensory processes, and thus might be less sensitive to inferential processes than magnitude estimation. This would explain the Libet Clock’s failure to detect binding at longer intervals. Moreover, it is well known that magnitude estimation is susceptible to a range of cognitive biases (e.g., see Poulton, 1979). Our results of binding at super-second intervals could thus equally reflect the operation of inferential processes, subject to postdictive biases.

In conclusion, we have expanded upon earlier demonstrations of intentional causal binding, and could show that the effect is robust when measured directly via verbal judgments, and stable across intervals as long as 4 s. Inferential, postdictive judgment biases and a slower pacemaker speed during anticipation of intentionally caused outcomes can both account for the effect. The trademark signature of different pacemaker speeds—reliable differences in slope—however suggests that pacemaker speed will have to considered in future research on temporal binding.

References


Received January 15, 2008
Revision received September 23, 2008
Accepted September 24, 2008