Consciousness and Cognition 20 (2011) 1518–1531

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Shifts of criteria or neural timing? The assumptions underlying timing perception studies

Kielan Yarrow [a,*], Nina Jahn [a], Szonya Durant [b], Derek H. Arnold [c]

[a] *Department of Psychology, City University London, United Kingdom*
[b] *Department of Psychology, Royal Holloway, University of London, United Kingdom*
[c] *School of Psychology, The University of Queensland, Australia*

ARTICLE INFO

Article history:
Received 27 August 2010
Available online 31 July 2011

Keywords:
Time
Simultaneity
Order
Perception
Recalibration

ABSTRACT

In timing perception studies, the timing of one event is usually manipulated relative to another, and participants are asked to judge if the two events were synchronous, or to judge which of the two events occurred first. Responses are analyzed to determine a measure of central tendency, which is taken as an estimate of the timing at which the two events are perceptually synchronous. When these estimates do not coincide with *physical* synchrony, it is often assumed that the sensory signals are asynchronous, as though the transfer of information concerning one input has been accelerated or decelerated relative to the other. Here we show that, while this is a viable interpretation, it is equally plausible that such effects are driven by shifts in the criteria used to differentiate simultaneous from asynchronous inputs. Our analyses expose important ambiguities concerning the interpretation of simultaneity judgement data, which have hitherto been underappreciated.

1. Introduction

Our ability to perceive the time at which events occur is an important topic of psychological investigation. Indeed, some have ascribed the birth of experimental psychology to an event at the Greenwich observatory in 1796, when Maskelyne, the astronomer royal, dismissed his assistant Kinnebrook for incompetence regarding the timing of transits, supposedly on the basis that the assistant's estimates differed systematically from his own (Mollon & Perkins, 1996). Although Kinnebrook's dismissal may have had more to do with his judgements about marriage than his judgements about transits (he had recently refused to accept a match suggested by his employer) the episode provoked another astronomer, Bessel, to investigate the so-called "personal equation": the differences in perceptual and reaction times across individuals. Wundt's school was to pursue this question in the latter half of the nineteenth century, and the general topic of temporal perception has remained active ever since.

1.1. Estimating perceived timing

The basic assumption underlying all studies of timing perception is that there might be a difference between physical and perceived event timings (an assumption that follows naturally from the distributed and delayed nature of neural activity). This entire field of enquiry is therefore inherently subjective. The two most common tasks used to assess perceived timing are simultaneity judgements (SJs, e.g. Fujisaki, Shimojo, Kashino, & Nishida, 2004; Moutoussis & Zeki, 1997a, 1997b; Stone

---

[*] Corresponding author. Address: Social Science Building, City University London, Northampton Square, London EC1V 0HB, United Kingdom. Fax: +44 (0)20 7040 8580.

E-mail address: kielan.yarrow.1@city.ac.uk (K. Yarrow).

1053-8100/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2011.07.003

et al., 2002) and temporal order judgements (TOJs; e.g. Eagleman & Sejnowski, 2000; McDonald, Teder-Salejarvi, Di Russo, & Hillyard, 2005; Sugita & Suzuki, 2003). In SJs, participants are shown a combination of two, or more, events and are asked whether or not they were synchronous. In TOJs, participants are asked to judge which event happened first.

In studies involving either TOJs or SJs, it is common to fit noisy data from individual participants with a continuous function to derive key parameters, such as a measure of central tendency. This is often taken as an estimate of the relative timing at which two events seem synchronous – the point of subjective simultaneity (PSS). This terminology itself is misleading, as there is seldom, if ever, a precise timing relationship at which two events seem synchronous and none other. Instead, there is typically a relatively broad range of timings at which events are at least sometimes judged as synchronous (see Fig. 1 parts A and B). However, this basic approach is almost ubiquitous in studies of timing perception, as is the tendency to assign meaning to the fitted parameters without fully discussing the assumptions underlying the continuous function fits, and the limitations that these place on interpretation. Indeed, we speculate that researchers might sometimes forget that by fitting a function and extracting one or more parameters, they are implicitly countenancing a model of the underlying psychological processes.

In what follows, we will consider some simple models that make use of concepts derived from signal detection theory (Green & Swets, 1966). The information-processing models we investigate may be considered rather abstract, in that they propose a continuous internal representation of stimulus magnitude (an internal response; in our case the time between two sensory events) without giving a detailed account of how this representation is generated or maintained in the brain. However, models specified at this level of abstraction have informed the curve fittings used in timing tasks (and indeed many other psychophysical judgements). A key feature of these models is that they distinguish between the information available, based on the internal response, and the interpretation of this information, based on the setting of one or more decision criteria. This is a distinction we will pursue here. Although we will sometimes allude to underlying neural processes in the course of developing our argument, we consider the precise neural instantiation of these models to be beyond the scope of this paper.

1.2. Assumptions underlying analyses of TOJs

We begin by considering the temporal order judgement. Here, two stimuli, *A* and *B*, are presented at a range of stimulus offset asynchronies (SOAs; *B*–*A*) and the observer judges their order. The proportion of times that stimulus *B* is judged to come second is usually an increasing function of SOA (see Fig. 1 part A). A straightforward model of this task assumes that each stimulus is accompanied by Gaussian noise affecting its central arrival latency, and that the two stimuli might be delayed by neural processing to different extents (Baron, 1969; Gibbon & Rutschmann, 1969). It is further assumed that these noisy and delayed arrival times produce a distribution of differences in central arrival time for any physical SOA, which is also Gaussian. This central arrival time can therefore be thought of as a kind of *subjective* SOA. Accordingly, the difference detected on any given trial must be compared to a criterion (zero for an unbiased observer) in order to reach a judgement. If the difference falls below the criterion, *B* is judged to have come first. If it falls above the criterion, *B* is judged to come second. This model, schematised in Fig. 2 part A, predicts a cumulative Gaussian psychometric function, which closely approximates the data observed in most timing studies.

Above, we have described perhaps the most basic model of the TOJ. There are a range of other TOJ models with similar basic assumptions, but more complex decision mechanisms. Each predicts a similar shaped function (Sternberg & Knoll, 1973). The slope of the function is taken to reflect both sensory noise and decision/criterion noise. The midpoint of the function is assumed to reflect any relative processing delay between the two stimuli (hereafter referred to as a *differential delay*, a value which can be added to the objective SOA to yield a subjective SOA, as depicted in Fig. 2) in addition to the placement of the criterion.

This last point is important. The PSS derived in this way from TOJ data *cannot unambiguously* indicate a differential delay for two stimuli, which is what we usually wish to infer, because a criterion shift is an equally plausible explanation of a shift in the measure of central tendency. For example, a PSS of −20 ms might reflect a differential delay of +10 ms and a criterion set at −30 ms, or any other combination summing to −20. It may seem strange that a criterion for judging arrival times should ever deviate from zero. Unfortunately, participants in psychophysical experiments seem rather prone to shifting their

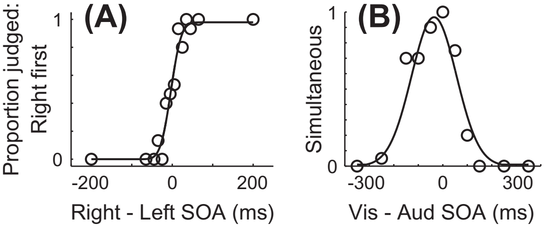

Fig. 1. (A) Illustrative data from a temporal order judgement experiment (auditory only, left versus right) fitted with a cumulative Gaussian curve. (B) Illustrative data from a simultaneity judgement experiment (visual-auditory) fitted with a Gaussian curve.

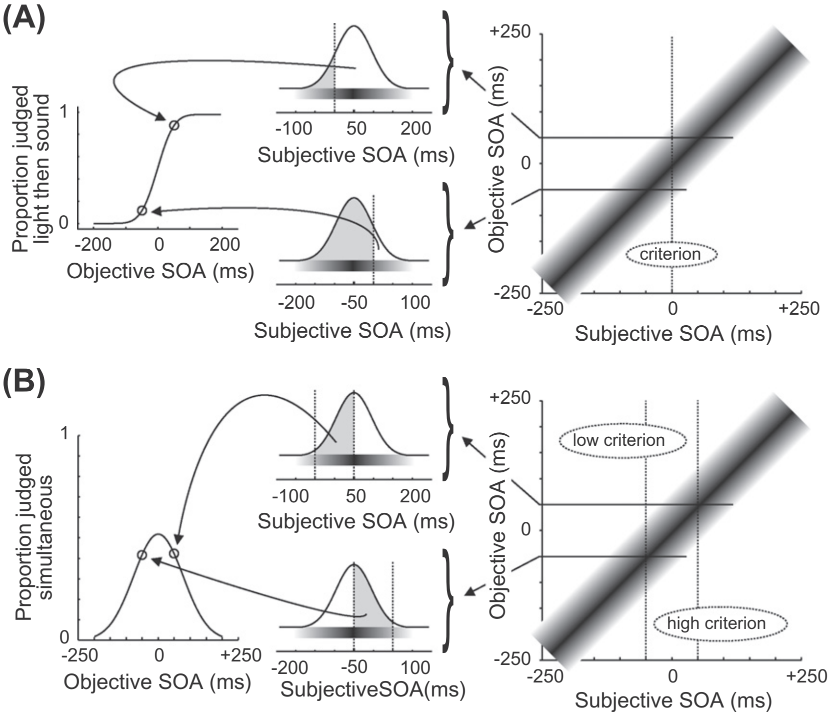

Fig. 2. (A) Schematic showing how the mapping between the objective SOA and an observer's subjective SOA (i.e. the differential delay) is combined with the placement of a criterion to give rise to a cumulative Gaussian psychometric function in a TOJ task. On the right, the function that relates objective and subjective SOAs is depicted. In this case, the relationship is veridical (i.e. a differential delay of zero), so subjective SOA will on average equal objective SOA. A differential delay, were it present, would result in a function that was displaced along the *x* axis. The presence of sensory noise is represented by the shading variations superimposed on the objective-to-subjective SOA function, with darker shading indicating higher probability density. This noise means that on any given trial, subjective SOA may not exactly equal objective SOA. The dashed vertical line indicates the observer's criterion (in this case unbiased, so set at zero) for choosing one of two responses: For subjective SOAs above the criterion, light is considered to precede sound, and vice versa for subjective SOAs below the criterion. Moving to the central column of the figure, two probability density functions are shown, each indicating how various subjective SOAs will be generated over repeated trials at just one of two example objective SOAs (−50 and 50 ms). These can be thought of as slices through the objective-to-subjective SOA function (from the right-hand graph). For each objective SOA, a proportion of the subjective SOAs will fall above the criterion (shown here as the unshaded region) and be judged "light then sound". These proportions are shown in the left-hand column, and yield a predicted psychometric function. (B) The same approach used in part A is taken to show how the differential delay, in combination with the placement of a two criteria, gives rise to a psychometric function equalling the difference of two cumulative Gaussians in an SJ task. Here, the observer classifies the stimuli "simultaneous" when the subjective SOA falls between a low criterion and a high criterion. Given sensory noise, for each objective SOA, sampled over multiple trials, a proportion of trials will be classified simultaneous (shaded area, central column). We can predict this proportion by measuring the distance from the left-hand side of the probability density function to the high criterion (i.e. a first cumulative Gaussian) and subtracting the distance from the left-hand side of the probability density function to the low criterion (a second cumulative Gaussian). This would produce the psychometric function shown on the left.

criteria based on incidental features of the question being asked (e.g. Bedell, Chung, Ogmen, & Patel, 2003; Clifford, Arnold, & Pearson, 2003), and it has proven rather difficult to completely eliminate these biases (Shore, Spence, & Klein, 2001; Spence, Shore, & Klein, 2001). This is one reason that some investigators have suggested, albeit tentatively, that simultaneity judgements might make a better choice of task than temporal order judgements when the PSS is an important parameter for investigation (Schneider & Bavelier, 2003; van Eijk, Kohlrausch, Juola, & van de Par, 2008; Zampini, Shore, & Spence, 2005).

1.3. Assumptions underlying analyses of SJs

Simultaneity judgements are similar to TOJs, except that participants judge whether two stimuli presented on each trial are presented at the same or at different times. This leads to a function which rises as the SOA approaches zero, then falls off again (see Fig. 1 part B). At first sight, the criterion used in SJs (i.e. a decision about when two events are close enough together to be judged synchronous) may appear less problematic than the one used for TOJs, because it seems to affect the height of the curve rather than any measure of central tendency. Interestingly, while the function commonly used to fit TOJs has a clear modelling rationale, the function most commonly used to fit SJs (the Gaussian or truncated Gaussian; e.g. Fujisaki et al., 2004; Stone et al., 2002) is rarely discussed, and seems to have been chosen primarily for convenience. This makes it rather difficult to spell out what the PSS derived from this function (usually taken as the peak of the fitted Gaussian) might represent.

To elucidate the often unstated assumptions underlying analyses of simultaneity data, we can take as a starting point the derivation provided by Schneider and Bavelier (2003). This derivation builds on work by Sternberg and Knoll (1973), Allan (1975) and Ulrich (1987). A slight variant has also recently been used to fit SJ data by Cravo, Claessens, and Baldo (2011). The model is as described above for the TOJ task, except that the observer is assumed to use two criteria rather than one. If the difference in central arrival times falls between these two criteria, the observer calls the stimuli simultaneous. This leads to a simultaneity function defined as the difference of two cumulative Gaussians. The model is schematised in Fig. 2 part B.

Leaving aside for the moment estimates of noise, the TOJ model we described earlier provides one parameter (the PSS) which reflects two degenerate variables (the differential delay and the placement of the criterion) while the SJ model provides two parameters (the midpoints of the two cumulative Gaussians) which reflect three degenerate variables (the differential delay and the extents of two criteria). To clarify the terminology we will use when referring to our SJ models and data: The *boundaries* are points along the objective timeline of SOAs, such that when the signals plus sensory noise fall between these two values the observer judges them as simultaneous. The *criteria* are decision boundaries applied to the *subjective* timeline of SOAs (depicted as dashed lines in Fig. 2 part B). The *extents* of the two criteria are the distances of each one from a subjective time zero, which may differ from objective time zero due to differential neural delays. Hence, if two cumulative Gaussians imply boundaries at −100 and +100 ms, for example, this could reflect a differential delay of +10 ms with a low criterion extent of 110 ms and a high criterion extent of 90 ms, or any other combination of values that add up correctly. This point bears reiteration: A fitted model of this kind tells us about boundaries, but the differential delay is free to trade off against the extent of the two criteria within these boundaries.

Hence for SJs, as for TOJs, it is only possible to estimate all of the underlying variables by making further assumptions. For example, in the more straightforward of their derivations, Schneider and Bavelier (2003) adopted the triggered-moment model (Venables, 1960). This model suggests that the first stimulus to arrive centrally triggers a time-consuming mental process, and that judgements of simultaneity occur when the second stimulus arrives while this process is being conducted. As arrival order does not matter in this account, the two criteria must be set at equal temporal distances relative to the coincidence of central arrival times (i.e. subjective time zero). If stimuli arrive centrally in the order *AB*, the observer says simultaneous if $B - A < X$ ms, whereas if they arrive in the order *BA*, they say simultaneous when $A - B < X$ ms. Under this model a PSS can therefore be derived which reflects only the differential delay, alongside a single criterion extent (*X*, which applies equally to both sides of the difference distribution). This PSS falls at the midpoint of the difference of cumulative Gaussians, so it would be similar to the PSS commonly obtained when a single Gaussian function is used to fit SJ data, but this model has the additional advantage of being able to capture a flattened peak due to a broad range of values being perceived as simultaneous.

One point we wish to make forcefully is that we do not think that the added assumption of the triggered moment model (i.e. equality in criterion extents) is particularly compelling. We will appeal to some established properties of the neural response to illustrate this point, although we recognise that the models we are describing are not fully specified at this level of analysis. Consider first that two events that are physically transient will give rise to internal neural activity that is persistent. Indeed, psychophysical findings suggest that sensory inputs undergo a form of low-pass filtering as they are processed in the brain (de Lange, 1958; Roufs & Blommaert, 1981; Viemeister, 1979; Weisenberger, 1986). Taking the example of an audiovisual pair, it is likely that the internal response to a visual stimulus is prolonged relative to the internal response to an auditory stimulus (de Lange, 1958; Viemeister, 1979).

This notion is illustrated in Fig. 3. We envisage that a temporal comparison between the two stimuli is achieved based upon the difference in their detection times, with detection being based on the rising portion of the impulse response. The difference in detection times gives us the subjective SOA. However, it does not tell us how a criterion is selected to determine whether the stimulus that is detected second is synchronous with the stimulus that is detected first. Given the situation depicted in Fig. 3, it seems quite reasonable that a more liberal criterion extent would be adopted when judging whether an auditory event which just followed a visual stimulus was simultaneous with it, compared to the reverse situation. We might think of this as something akin to applying the rule: “If the second stimulus arrives while I can still perceive the first stimulus, they are simultaneous”. This would imply independent criteria for the two orders of stimulus presentation, be-

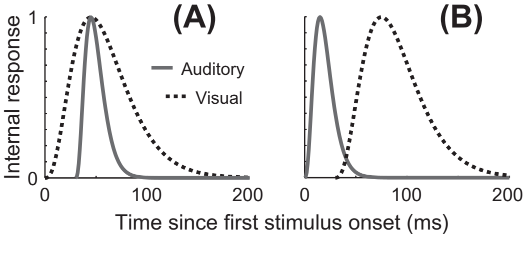

Fig. 3. (A) Hypothetical neural activity (impulse responses) in reaction to a brief visual event followed after 30 ms by a brief auditory event. The filter applied to the visual response has a longer time constant. In this example it seems reasonable that a criterion could be placed some way into the visual response such that the auditory response would be detected before this point and thus judged simultaneous. (B) Identical internal activity with the order of visual and auditory responses reversed (i.e. the visual event follows the auditory event by 30 ms). Here it seems unlikely that the visual response would be detected prior to reaching a criterion placed anywhere along the auditory response. It would therefore be unlikely to be judged simultaneous.

cause lights persist longer than sounds. Furthermore, it seems quite reasonable that a particular experimental context or instruction could give rise to a shift in one criterion that is not automatically mirrored by a shift in the other.

If independent criteria are in place, one must be careful when interpreting the PSS from SJ tasks, just as one must be careful in interpreting the PSS from TOJs. We can only say where two boundaries are positioned relative to the objective SOA axis, and that the differential delay should fall between these boundaries. These issues are illustrated in Figs. 4 and 5, which show the psychophysical functions that are predicted to arise under the two-criterion SJ model when criteria and differential delays are varied. The main point to note from Fig. 4 is that an identical change between the data observed in two experimental conditions can, in principal, arise from either a change in differential delay or from a change in the placement of criteria. However, Occam's razor can be used to provide some support for one interpretation over the other. The main point to note from Fig. 5 is that even when employing the law of parsimony to aid interpretation, the difference between data that would favour a change in the differential delay and data that would favour a shift in a criterion is likely to be rather subtle.

The critical issue at hand, then, is whether data which are *consistent* with subjective timing shifts can be taken as unequivocal evidence for a shift in the relative time course of neural responses for two or more inputs. Certainly it is possible that a PSS shift, as indicated by a measure of central tendency in a continuous function fitted to TOJ or SJ data, reflects an altered time course of neural activity. However, we will argue that it is at least equally plausible that the time courses of neural activity remain unchanged, but that criteria concerning temporal judgments are altered.

In what follows we present two analyses of temporal judgement data. We take as a starting point the model outlined above and use it to evaluate the possibility of a neural timing shift in different experimental conditions. In the first analysis we adapt an approach introduced by Allan (1975) and developed by Ulrich (1987) to suggest that the SJ model requires addi-

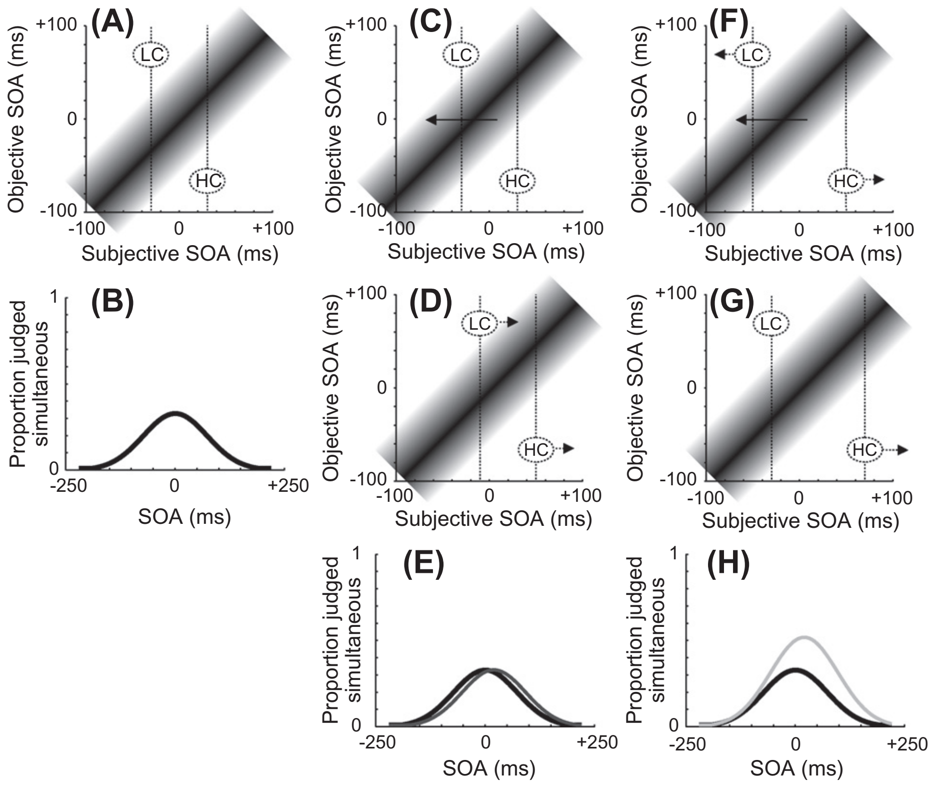

Fig. 4. (A and B) Schematic showing how the differential delay (represented by the intercept of the linear function in part A) and the placement of two criteria (represented by dashed vertical lines: LC and HC = low/high criteria respectively) give rise to particular psychometric functions under a two-criterion SJ model (see legend to Fig. 2 for additional explanation). As can be seen from the exact correspondence between objective SOA and the most likely subjective SOA, this observer is veridical (i.e. differential delay = 0). They are also unbiased (i.e. equal low and high criterion extents). Part B shows the data function predicted by a two-criterion SJ model for the observer described in part A. (C–E) Part E shows the same function as part B (from a veridical and unbiased observer, in black) alongside a dark grey function that simulates possible responses after some experimental manipulation has affected measures of central tendency. The dark grey function is shifted entirely to the right. Possible causes for such a shift are schematised above in parts C and D. The most parsimonious explanation of such a shift would be that the differential delay for this observer has moved towards a positive PSS (shown in part C by a leftward shift of the objective–subjective SOA function). An alternative explanation would be that the low criterion has tightened up while the high criterion has loosened, with no change in the differential delay (shown in D). (F–H) The same format is used as in panels C, D and E. Panel H shows veridical and unbiased performance (black) alongside new data obtained following some experimental manipulation (light grey function). The shape of the light grey function has changed: It has expanded, with a shift in central tendency to the right. Possible causes are shown in parts F and G. The most parsimonious explanation is that only the high criterion has changed, becoming looser (shown in G). An alternative explanation would be that the differential delay has moved a small amount towards a positive PSS while the criteria have loosened on both sides (shown in F).

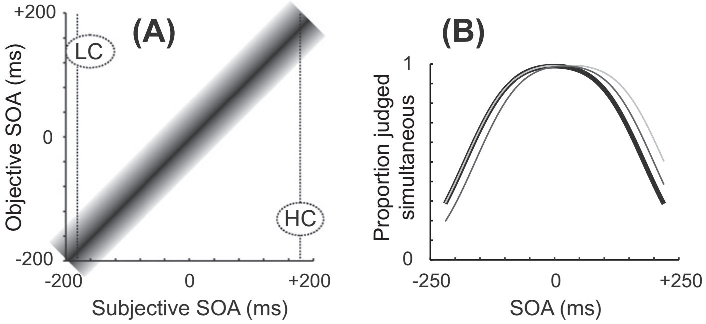

Fig. 5. (A) Schematic showing how the placement of two criteria (represented by dashed vertical lines: LC and HC = low/high criteria respectively) give rise to particular psychometric functions under a two-criterion SJ model. The format is identical to that used in Fig. 4, which depicted an observer employing tight criteria (i.e. a small criterion extent) relative to the magnitude of their sensory noise in order to emphasise differences in the shape of the psychometric function when criteria relax. Here, a comparable representation is provided for an observer employing much looser criteria (i.e. a larger criterion extent) but with identical sensory noise. The shape of these data is more typical of audiovisual SJs like those used in the experiment we describe here. (B) The black (veridical, unbiased) function represents data from the observer depicted in panel A. The two grey functions simulate possible responses after some experimental manipulation has affected measures of central tendency. Based on parsimony (for more details see the legend to Fig. 4) the dark grey function (shifted entirely to the right) would be explained most simply by a shift in the differential delay towards a positive PSS. The light grey function (which has expanded, with a shift in central tendency to the right) would be explained most simply by a loosening of the high criterion. Note that with these loose criteria, which are more representative of typical performance than those illustrated in Fig. 4, the differences between the two situations becomes much less striking (c.f. Fig. 4 panels E versus H).

tional noise in the criteria to better capture the data. Note that this analysis does not relate to the question of how a shift in central tendency should be interpreted, but rather helps us to determine the simplest model that will adequately capture our data. In the second analysis we apply the resultant extended model, incorporating criterion noise, to our interpretative question, and find that criterion shifts provide as good, or an even better, account of the observed data relative to an assumed shift in neural timing.

1.4. Our SJ dataset – temporal recalibration

The dataset we have chosen for analysis concerns a phenomenon known as temporal recalibration (Fujisaki et al., 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004). We will discuss this phenomenon in more detail soon, but wish to point out from the outset that the issues at hand are not specific to this one timing phenomenon. Rather, they are common across every instance wherein experimenters have inferred a shift in neural processing times on the basis of an apparent PSS shift in different experimental conditions (e.g. colour-motion asynchrony, Moutoussis & Zeki, 1997a; audio-visual synchrony with viewing distance, Sugita & Suzuki, 2003; prior entry, Schneider & Bavelier, 2003). However, note that we are not claiming that the particular interpretation we reach here (regarding the relative contribution of differential delays and the placement of criteria) will generalise to all, or indeed to any, of these other cases. Rather, we are simply pointing out that similar interpretative ambiguities arise in any SJ data set. This fact seems to have been acknowledged by many authors for TOJs (e.g. Shore et al., 2001) but perhaps not for SJs (e.g. van Eijk et al., 2008).

In temporal recalibration experiments (Fujisaki et al., 2004; Vroomen et al., 2004), participants are exposed to long sequences of paired adapting events, consisting for example of beeps and flashes. They are then given intermittent test trials to establish the point at which two subsequent events seem simultaneous (the point of subjective simultaneity or PSS). When beeps *lag* flashes during adaptation, the PSS tends to shift in the same direction, so that a beep lagging a flash is now judged as more simultaneous. A reverse effect can ensue following adaptation to a stimulus wherein beeps *lead* flashes.

Temporal recalibration has been assessed with a variety of tasks, including simple reaction times (Di Luca, Machulla, & Ernst, 2009; Harrar & Harris, 2008; Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009) and the perception of ambiguous visual stimuli in bimodal displays (Fujisaki et al., 2004). However, it is typically measured as a shift in the PSS derived from TOJs or SJs between conditions with different adapting asynchronies. Our primary concern here is that temporal recalibration is a good example of a phenomenon wherein a PSS shift across experimental conditions has been taken as evidence for altered neural processing times.

2. Methods

2.1. Participants

We tested 19 participants (five male, mean age = 21.8, SD = 2.5) who took part in the experiment for a monetary reward. One participant had to be rejected because no reliable fit could be obtained in one or more conditions (see data analysis, below). This yielded a sample size of 18 (five male, mean age = 21.8, SD = 2.5). The study was approved by the City University London psychology department's research ethics committee.

2.2. Apparatus and stimuli

The experiment was controlled by a PC sending digitised signals at 44,100 Hz using a 12 bit A/D card (National Instruments DAQCard 6715). We confirmed the correct timing of output signals using a 20 MHz storage oscilloscope (Gould DSO 1604). Participants sat at a comfortable distance (~50 cm) from a computer monitor. A small speaker and a red light-emitting diode (LED) mounted in a vertical backing were positioned centrally and just in front of the monitor. Visual stimuli were 10 ms LED flashes (or 100 ms in the case of infrequent oddball stimuli presented during adaptation). Auditory stimuli were 10 ms 1000 Hz pure tones (or 100 ms in the case of oddball stimuli).

2.3. Procedure

Each block of trials began with an adaptation phase containing 200 (~1 Hz) presentations of audio-visual adaptation stimuli. In separate blocks (with a fixed order) adapting tones and flashes were either presented in physical synchrony (0 ms adaptors), with tones preceding flashes by 150 ms (−150 ms adaptors), or with flashes preceding tones by 150 ms (+150 ms adaptors). Also in separate blocks (with a counterbalanced order) adaptation stimuli were either presented at fixed times after the participant had pressed a mouse button, at variable times after the participant had pressed a mouse button, or while the participant passively viewed the display. Because this manipulation yielded no reliable main effects or interactions in a preliminary analysis, data from these latter conditions were pooled in order to improve the reliability of model fits. Each category of block was presented twice in succession, yielding a total of eighteen blocks of trials for each participant.

To ensure that participants attended to adapting stimuli, we included a secondary task. Adaptors were 90% standards (10 ms long tone and flash), 5% visual oddballs (10 ms tone, 100 ms flash) and 5% auditory oddballs (100 ms tone, 10 ms flash). Participants were required to say “oddball” whenever a long stimulus was detected, and this response was noted by the experimenter along with the adaptation trial (displayed for this purpose in the top left corner of the computer screen).

Following adaptation (or top-up, see below), the entire screen turned red for 2 s to signal that a test trial was approaching. After a further pseudorandom delay of 1000–1500 ms, the LED flashed once. An auditory tone was also presented around this time, with an SOA of −450 to +450 ms relative to the flash. This SOA was varied from trial to trial. It was selected randomly from a condition-specific distribution. Each distribution was initially uniform, containing delay values from −210 to +210 ms in 30 ms increments, but was updated after each accepted trial according to the generalised Pólya urn model (Rosenberger & Grill, 1997) based on judgements made about temporal order (see below). Distributions could therefore expand to include delay values from −450 to +450 ms. This procedure produces a majority of values close to the point of subjective simultaneity.

After each test stimulus presentation, a response window appeared on the screen and participants were invited to first complete an SJ (“did the flash and the beep seem like a single event?”) and then a TOJ (“which came first, the beep or the flash?”). A cancel trial option was available in case the participant had been distracted, in which case the trial was added to the end of the block. Following their judgements, an on-screen message told participants to press the mouse button to initiate a top-up phase of adaptation. This was identical to the original adaptation phase, except that it contained only eight adaptors. Each block terminated after 35 test trials had been logged, generating 210 trials per adapting asynchrony condition.

2.4. Data analysis

Our first analysis was intended to test the sufficiency of the two-criterion SJ model outlined in the introduction. As first noted by Allan (1975) and developed by Ulrich (1987), the use of a third “simultaneous” category, in addition to the two order responses of a classic TOJ, permits the generation of two psychometric functions. One of the two psychometric functions is displaced along the SOA axis relative to the other. The relationship between these two functions can permit inferences about the judgement process – for example, about whether a two-criterion model is plausible. In fact, under the two-criterion SJ model, these functions should be identical to the two cumulative Gaussians from which a difference of cumulative Gaussian fit to SJ data can be obtained (see Schneider & Bavelier, 2003, Appendix A.1 for a derivation).

For the current data, the leftmost function was estimated using the probability (p) of judging the stimuli as both successive (Su) and in the order sound then light (SL). An increasing function was required, and was calculated as $1 - p(\mathrm{SL} \cap \mathrm{Su})$. This function relates to the boundary between judging that the sound preceded the light and that the two stimuli were simultaneous, i.e. the low boundary. The rightmost function was estimated as $p(\mathrm{LS} \cap \mathrm{Su})$. This function relates to the boundary between judging stimuli as simultaneous and judging that the light preceded the sound, i.e. the high boundary. The two-criterion SJ model implies that these two functions should be cumulative Gaussians with identical slopes, but displaced intercepts (Allan, 1975). However, a range of other relationships are possible if we relax the model’s assumptions (Ulrich, 1987). We therefore fitted these joint probabilities with cumulative Gaussians using the psignifit toolbox version 2.5.6 for Matlab (see http://bootstrap-software.org/psignifit/). This implements a maximum-likelihood method, as described by Wichmann and Hill (2001). Judgement uncertainty was estimated as the difference between the SOA values that yielded probabilities of 0.5 and 0.84. We rejected one participant because the range of SOAs tested was insufficient to form a reasonable estimate of both the low and high boundaries (the data included only a single trial more extreme than the estimated boundary for three out of six estimates).

As outlined in Section 3, this first analysis suggested that we should permit noise in the two boundaries/criteria in order to best fit these data. This allows for the position of each criterion to shift from trial to trial, rather than remaining fixed. Our second analysis therefore expanded our initial two-criterion SJ model to fit a function of the form:

$$p\ \text{“simultaneous”} = \Phi(B_{\text{High}}, \Delta t, \sigma_{\text{High}}) - \Phi(B_{\text{Low}}, \Delta t, \sigma_{\text{Low}}) \tag{1}$$

where $\Phi(B_{\text{Low}}, \Delta t, \sigma_{\text{Low}})$ denotes a cumulative Gaussian (identical to the integral of a Gaussian distribution from negative infinity to point B_{Low}, the low boundary), with parameters Δt (i.e. the SOA) and σ_{Low}. This final parameter represents:

$$\sigma_{\text{Low}} = \sqrt{}(\sigma^2_{\Delta D} + \sigma^2_{\text{BLow}}) \tag{2}$$

where $\sigma^2_{\Delta D}$ is the variance of the difference distribution for central arrival times of flashes and beeps (i.e. $\sigma^2_{\text{Flash}} + \sigma^2_{\text{Beep}}$) and σ^2_{BLow} is the Gaussian trial-to-trial variance in the placement of the low boundary (or equivalently, the low criterion). Note that the first part of Eq. (1) can be unpacked in identical fashion, replacing references to the low boundary with references to the high boundary.

In order to gain maximum benefit from our dual SJ and TOJ questions, we fitted simultaneity data using Eq. (1), whilst simultaneously fitting the leftmost and rightmost functions (based only on TOJs for trials judged successive, as outlined for analysis one) to the component cumulative Gaussians of the simultaneity fit (the low boundary and high boundary components respectively). Hence the same two curves had to fit both the order data (from successive trials) and the simultaneity data. This approach makes full use of a ternary division of responses (sound-first/simultaneous/light-first), ensuring that the fit of the model will be made worse when trials judged successive are then assigned the wrong order. A maximum-likelihood fit was obtained using the Nelder–Mead simplex algorithm (Nelder & Mead, 1965; O'Neill, 1971). The fit yielded four parameters: The two boundaries (which can be averaged to provide an estimate of differential delay under the triggered moment model, or alternatively interpreted as limits on the differential delay in our preferred interpretative scheme) and two measures of judgement variability, σ_{Low} and σ_{High}.

The two derived measures of judgment variability (σ_{Low} and σ_{High}) reflect three sources of noise: Sensory noise associated with the incoming stimuli, and two further sources of noise reflecting variability in the placement of the low and high criteria. We cannot uniquely estimate all three sources, but we can derive two useful measures from our parameters. First, we can estimate an upper limit on the sensory noise (equal to the smaller of σ_{Low} and σ_{High}). Second, we can estimate the difference in trial-to-trial variance between high and low criteria (equal to $\sigma^2_{\text{High}} - \sigma^2_{\text{Low}}$). Here, a positive value indicates that participants are more variable in the placement of their criterion when sounds follow lights than vice versa.

To complete our second analysis, we recorded the deviance of the model fit for each participant and condition, and then performed a second fit in an identical manner, but using a simpler model without noisy criteria. Deviance is an appropriate measure of goodness of fit for maximum-likelihood fits (Wichmann & Hill, 2001). Because the two models are nested, we can predict how much deviance should improve with the addition of a single parameter (it should follow a chi-squared distribution with one degree of freedom; Wichmann & Hill, 2001). This provided an extra test regarding the appropriateness of the more complex model.

3. Results

3.1. Analysis one

We recorded both TOJs and SJs, on every test trial. Combining TOJs and SJs permits us to measure two functions[1] (Allan, 1975; Ulrich, 1987). The leftmost function relates to the boundary between judging that the sound preceded the light and judging that the two stimuli were simultaneous, while the rightmost function relates to the boundary between judging stimuli as simultaneous and judging that the light preceded the sound.

Fig. 6 parts A–C show examples of these functions for one participant. The two-criterion SJ model suggests that both functions should be cumulative Gaussians of identical slope (Allan, 1975). However, this model forms part of a larger class of model termed "general threshold models" (Ulrich, 1987, pp. 225) which share the assumption of noise in perceptual latencies but make less restrictive assumptions about the shape of noise distributions and, in particular, the presence or absence of criterion noise. For simplicity, we retained the notion of Gaussian noise and fitted our leftmost and rightmost function with cumulative Gaussians. As previously noted, noiseless (i.e. identical from trial to trial) criteria imply parallel functions, because sensory noise is present in equal measure regardless of the placement of the criterion. However, differential noise in the two criteria can give rise to functions with different slopes (Ulrich, 1987). We therefore obtained a measure of judgement uncertainty from our fitted functions (proportional to the inverse slope) to assess the need to include criterion noise.

---

[1] In fact, three functions can be derived: The leftmost and rightmost functions we describe here, which quantify TOJs on trials judged successive, and the classic TOJ function that quantifies TOJs on all trials (which would be expected to lie somewhere between the leftmost and rightmost functions). We have not made use of the classic function because many of our participants opted not to guess temporal order on trials they judged as simultaneous, instead making a consistent response that rendered the classic TOJ function almost identical to one or other of the leftmost and rightmost functions. This means that the classic function has little additional capacity to discriminate between different models. TOJ data on trials judged simultaneous have therefore essentially been discarded and are not treated in any of our analyses, making our data set identical to that obtained from a ternary response (i.e. sound first/simultaneous/light first).

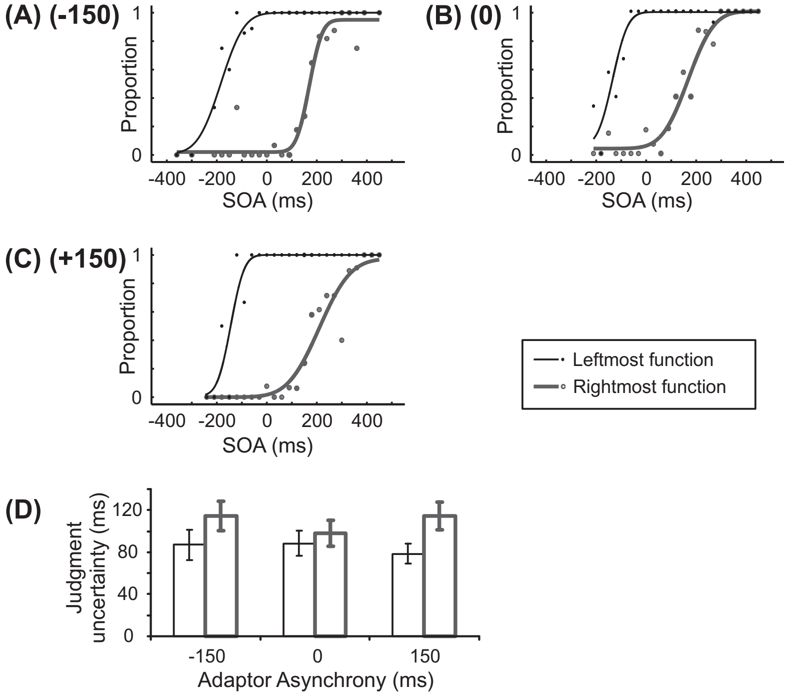

Fig. 6. Results from analysis one. (A–C) Fits of leftmost and rightmost functions for one example participant. Leftmost function = 1 − proportion(sound-then-light ∩ successive). Rightmost function = proportion(light-then-sound ∩ successive). (D) Group means derived from individual fits for judgement uncertainty (inverse slope) at each adaptor asynchrony. Error bars show standard error of the mean.

Fig. 6 part D shows the mean judgement uncertainty across participants for leftmost and rightmost functions at each adaptor asynchrony. The rightmost function shows consistently higher judgement uncertainty (i.e. a consistently shallower slope) than the leftmost function. These data were subjected to a 2 (function: leftmost versus rightmost) × 3 (adaptation: −150, 0, +150 ms) repeated-measures ANOVA, which revealed a main effect of function ($F = 5.3$, df = 1, 18, $p = .034$; leftmost mean = 88 ms, rightmost mean = 110 ms). There were no other reliable effects, although the interaction approached significance ($F = 3.1$, df = 1.8, 33.0, $p = .065$) suggesting differences may have been greater following −150 and +150 ms adaptation. It seems that participants may have had difficulty maintaining a consistent strategy for judging when simultaneity gave way to successiveness across trials.

3.2. Analysis two

Because the leftmost and rightmost functions differed reliably in our first analysis, we developed the two-criterion SJ model into a two-noisy-criterion SJ model, with the addition of separate sources of Gaussian trial-to-trial noise in the placement of the two criteria. We fitted this model to each participant's data for each adaptor asynchrony. We also performed an identical fit for the simpler two-criterion SJ model. We then recorded the deviance of the fit for the two-noisy-criterion SJ model and compared it with the deviance of the fit (to the same data sets) for the simpler two-criterion SJ model. If the models have similar predictive power we should expect the decrease in deviance to exceed 3.84 only 5% of the time. The mean decrease in deviance was 9.11, 6.69 and 8.49 for the −150, 0 and 150 ms adaptor asynchronies respectively, with 34/54 individual fits exceeding the critical 5% value. This reinforces our first analysis, suggesting that an extra term included to help capture criterion noise was a justified addition to the original model.

With this finding in mind, we adopted the two-noisy-criterion model in order to assess how performance changed across the three adaptation conditions. Fig. 7 parts A–C shows example data and fits for one participant (the same participant whose data are shown in Fig. 6). The derived dependent variables of interest are the positions of the two boundaries (Fig. 7 parts D and E), the upper limit on sensory noise (Fig. 7 part F), and the difference in criterion variance (Fig. 7 part G).

The low boundary represents the objective SOA at which participants changed from sound-first responses to simultaneous responses. Not surprisingly, it is generally negative (i.e. at an SOA where sounds preceded lights). It shifted further in a negative direction following −150 ms adaptation (where adapting sounds preceded lights) but was similar for 0 ms (synchronous adaptation) and +150 ms adaptors (where adapting sounds followed lights). A one-way repeated-measures ANOVA showed a significant effect of adaptor asynchrony ($F = 14.4$, df = 1.6, 27.8, $p < .001$), with follow-up tests showing significant differences between the −150 ms condition and both the 0 ms ($p < .001$) and +150 ms ($p < .001$) conditions. In contrast, the high boundary (where participants partitioned simultaneous and flash-first responses) was generally positive, and became more positive following +150 ms adaptation, being similar for 0 ms and −150 ms adaptors. Again, an ANOVA showed a significant effect of adaptor asynchrony ($F = 3.4$, df = 1.9, 32.9, $p = .045$), with follow-up tests this time showing significant differences between the +150 ms condition and the −150 ms condition conditions ($p = .012$), although the difference between the +150 ms condition and the 0 ms condition only approached significance ($p = .099$).

We have presented our data in terms of the placement of two boundaries. Under this analysis it looks as though the criterion for judging simultaneity was relaxed (i.e. shifted to larger absolute SOAs) exclusively on trials where stimuli arrive in the same order as that experienced during concurrent adaptation. This is the interpretation we favour. A more typical way to plot these data would be to adopt additional assumptions (either explicitly, e.g. based on the triggered moment model, or otherwise) and thus present a PSS at the midpoint of the two boundaries. A quick calculation tells us that in this case the PSS would shift in the direction of the adapting asynchrony, as is commonly found in studies of temporal recalibration (PSS = 10, 24 and 33 ms for −150, 0 and +150 ms adaptor asynchronies respectively). For completeness, the mean criterion extent (relative to the PSS) should also be reported in such an analysis. Here it would grow slightly following −150 ms and +150 ms adaptation compared to the 0 ms condition (criterion extent = 132, 123 and 129 ms for −150, 0 and +150 ms adaptor asynchronies respectively). Hence one interpretation of the data presented in Fig. 7 is that adaptation has led to a change in differential neural delays, along with the adoption, consciously or otherwise, of a looser criterion for simultaneity. Indeed, there are a number of ways in which a change in the differential delay alongside shifts in one or both criteria might yield our result. However, the single criterion shift interpretation is more parsimonious, as we will develop in our discussion.

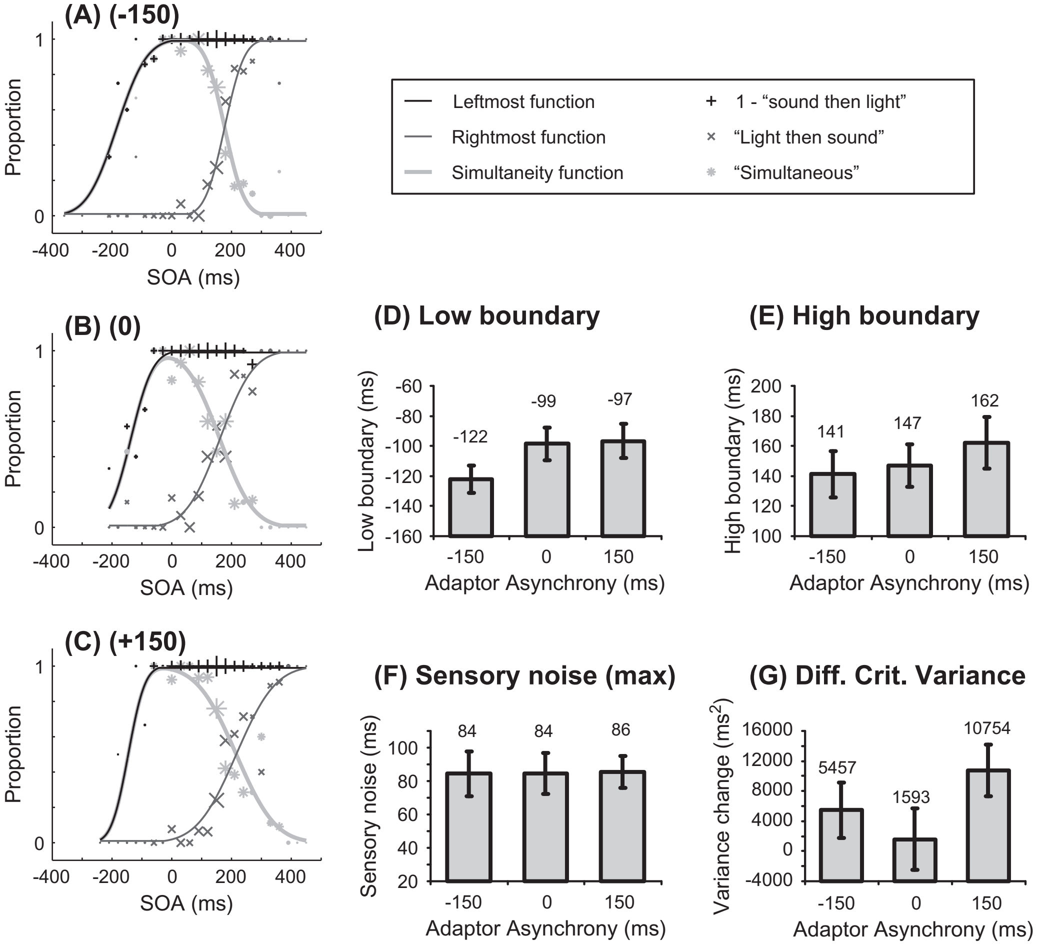

Fig. 7. Results from analysis two. (A–C) Fits of leftmost, rightmost and simultaneity functions for the same example participant shown in Fig. 6. The size of individual data points reflects the number of trials presented at each SOA. Leftmost function data points = 1 − proportion(sound-then-light ∩ successive). Rightmost function data points = proportion(light-then-sound ∩ successive). Simultaneity function data points = proportion(simultaneous). Leftmost function = low boundary cumulative Gaussian. Rightmost function = high boundary cumulative Gaussian. Simultaneity function = High boundary cumulative Gaussian − low boundary cumulative Gaussian. (D–G) Group means derived from individual fits for low boundary (D), high boundary (E), upper limit on sensory noise (F) and difference in criterion variance (G) at each adaptor asynchrony. Error bars show standard error of the mean.

We also looked at sensory and criterion noise. In this study sensory noise would not be expected to vary between adaptation conditions, and indeed the obtained values, which represent an upper limit on sensory noise, were very similar to one another with no reliable difference emerging (Fig. 7 part F). Criterion noise could not be estimated separately for each criterion, but the difference in criterion variance between high and low criteria was available (Fig. 7 part G). The generally positive values suggest that the high criterion may have been positioned with greater variability from trial to trial than the low criterion (i.e. the criterion shifted around more when sounds followed lights than vice versa). The difference across conditions approached significance ($F = 3.1$, df = 1.6, 27.3, $p = .069$). This trend suggests that the high criterion may have shown greater variability relative to the low criterion following adaptation, particularly following positive adaptation (when the mean position of the criterion had also moved outwards).

4. Discussion

Here we have examined a phenomenon, audio-visual temporal recalibration (Fujisaki et al., 2004; Vroomen et al., 2004), wherein adaptation to asynchronous audiovisual inputs alters subsequent judgments of audiovisual timing. Our data suggest that when people are subjected to audiovisual adaptors with a particular temporal order, the criterion that demarcates judgements of simultaneity from judgements of successiveness for that same stimulus order is relaxed, shifting outward toward the adapted timing offset. For example, when exposed to light followed by sound, participants were more likely to judge light-then-sound pairs as simultaneous, but showed no tendency to judge sound-then-light pairs as *less* simultaneous than before. We also noted a tendency for this effect to be greater when adapting to sounds that lagged lights than when adapting to lights that lagged sounds.

4.1. Realignment of modality timelines versus shifts in criteria

Our data are consistent with a shift in just one of two timing criteria following adaptation to asynchronous audiovisual inputs, specifically the criterion that generates judgements when stimuli arrive in the same order as the adaptors. We note that our data look rather similar to the group average data presented in one of the two original reports of temporal recalibration (Fujisaki et al., 2004). In that paper, and subsequent temporal recalibration studies, analyses were restricted to a consideration of the central tendencies of apparent audiovisual timing distributions. Here we have conducted more extensive analyses, establishing that such data can be accounted for via a model that assumes the existence of two timing criteria, of which only one is shifted post adaptation.

Unfortunately, on the basis of our data, it was impossible to unequivocally distinguish between a selective shift of just one timing criterion (as schematised in the change from Fig. 4 part A to Fig. 4 part G) and a change in the differential delay between vision and audition that is accompanied by a *smaller* outward shift of both criteria (i.e. a true realignment of sensory timelines *plus* a small symmetric relaxation of timing criteria; Fig. 4 part A versus F). This discrimination would only be possible if, post adaptation, one of the two boundaries (the transition points between perceived synchrony and asynchrony) were to shift *past* the unadapted position of the other boundary. For instance, in our study this could have happened if, post adaptation, the timing at which perceived successiveness (vision-then-audition) gave way to synchrony was shifted beyond the pre-adaptation timing at which perceived synchrony had given way to audition-then-vision successiveness. Note that this ambiguity in interpretation for SJ data is a central message that we wish to convey in this paper.

We prefer to interpret our results as showing that the adaptation-dependent shifts we measured were confined largely to one of the two criteria, rather than encompassing a differential delay, for the following reasons. First, this account is more parsimonious: An equivalent account, encompassing a realignment of neural signals, would require that the realignment be accompanied by a symmetric relaxation of timing criteria (see Fig. 4). Second, as outlined in Fig. 3, the temporally-smeared neural representations of transient sensory events dictate that criterion shifts are plausible. Third, criterion shifts along a low-pass filtered neural response imply a fairly tight limit on recalibration magnitude, as the criterion cannot reasonably be placed outside the boundaries of the neural response. This is consistent with the fact that temporal recalibration has never been found to approach the magnitude of the adapting asynchrony, but is rather much smaller. Fourth, temporal recalibration is almost invariably tested in a blocked design, which is likely to give scope for criterion shifts. Finally, seemingly high-level factors, such as the allocation of attention, can affect recalibration magnitude (Heron, Roach, Whitaker, & Hanson, 2010).

Of course, none of the arguments described above are decisive. Against them must be weighed the following consideration at least: Recalibration can be observed using tasks other than simultaneity judgements, including temporal order judgements, simple reaction times, and the stream-bounce illusion (e.g. Di Luca et al., 2009; Fujisaki et al., 2004; Hanson, Heron, & Whitaker, 2008; Harrar & Harris, 2008; Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009). However, all of these tasks also involve criterion settings, even the apparently implicit judgements in the multimodal bounce-stream illusion (Grove, Ashton, Kawachi, & Sakurai, 2009).

One should also keep in mind that the interpretations we have discussed are only valid in the context of the model we have been considering (which is examined in more detail below). However, we consider this an improvement over drawing conclusions without any model at all. It is likely that by making the model more complex, we could have given it a greater capacity to explain the data via either differential delays or criterion shifts. For example, we have considered only a simple additive transformation between objective SOAs and subjective SOAs in order to model a differential delay. If we permitted a non-linearity in this transformation (e.g. a gain below 1.0 applied only to SOAs on the adapted side, so that they seem subjectively decreased in magnitude), it might provide an alternative account of the recalibration we observed without recourse to changes in the criterion.[2] However, we would first want to demonstrate that such an increase in model complexity was justified by some feature of the experimental data or physiology.

A final interpretative issue arises because of the specific task we have used in this study. We used a combination of an SJ and a TOJ, one after the other, which we have effectively transformed into a ternary judgement for analysis (by ignoring TOJ data on trials judged simultaneous). It is possible that when participants must keep two questions in mind, they behave in a different way to when they must judge only one thing (see Allan, 1975, for a direct comparison). Hence it might be interesting to apply our analysis to a data set where only SJs were made, to check whether the same pattern emerges (we suspect that it would). Many of our participants seemed to translate the two tasks into a single ternary judgement (evidenced by a very consistent order judgement on trials judged simultaneous), but some participants did attempt to indicate order within the simultaneity zone. We note that the model we use here would have no problem dealing with above-chance performance in the simultaneity zone (because different criteria may be used for SJs versus TOJs). Such data would only be paradoxical under a more constrained model in which the SJ criteria are taken to reflect some kind of hard limitation, such as the time spent processing the first signal (Allan, 1975).

4.2. Two noisy criteria for judging simultaneity

Our modelling began with a version of the general independent channels model (Sternberg & Knoll, 1973) employing a deterministic decision rule (Baron, 1969; Gibbon & Rutschmann, 1969) based on the placement of two criteria for judging simultaneity and temporal order (Allan, 1975). Our first analysis found this model to be wanting. The slope of the leftwards and rightwards psychometric functions differed, implying a simultaneity function that is steeper on one side than on the other. This is in line with data from earlier studies (summarised in Ulrich, 1987) and with more recent research (van Eijk et al., 2008). We therefore turned to a less constrained version of the independent channels model, the general threshold model (Ulrich, 1987) and selected a variant which retained Gaussian sensory noise but also permitted Gaussian noise in the trial-by-trial placement of criteria. Our fits suggested that this model did a better job of capturing our data.

We have commented at length on the issues that arise when attempting to infer a differential delay based on SJ data, noting in particular that the SJ faces some of the same interpretative issues as the TOJ when it comes to distinguishing changes in neural timing from adjustments of criteria. It therefore seems appropriate to make some parallel comments about inferences drawn based on the *slopes* of fitted functions. Intuitively, it seems as though the SJ task yields a function that could vary in shape because of either differences in sensory noise or the adoption of more or less relaxed criteria. The temptation might therefore be to assign preference to the TOJ task when estimating variability, as it may appear to be less affected by response criteria. However, our modelling makes clear that it is actually quite straightforward to distinguish the extent of the criteria from the presence of noise in the SJ task. There is a more fundamental problem, however, which is once again the same for both tasks: It is very difficult to separate sensory noise from noise in the placement of the criterion/criteria.

The modelling we applied has the benefit that it makes explicit where the difference in slopes on either side of the simultaneity function might arise (i.e. from criterion noise) as well as retaining a perspective on the potential trade-off between differential delays and the extent of high and low criteria. Of course, other models may well be able to account for our data (e.g. Allik & Pulver, 1994; Burr, Silva, Cicchini, Banks, & Morrone, 2009; Kristofferson, 1967a, 1967b; Miller & Schwarz, 2006; Stelmach & Herdman, 1991). We should also note that the model we describe may require elaboration to deal with extant data sets, such as the tendency for high and low criteria to be drawn from distributions with negative and positive skews respectively (Ulrich, 1987) and the apparent dissociation of order and simultaneity judgements when attention is focussed on just one of two visual events (Stelmach & Herdman, 1991).

Our model is fairly straightforward when considered within the conceptual framework of signal detection theory, which emphasises the distinction between the information available in the internal response, and the criteria used to interpret that information. However, a potential limitation of our model is that the precise meaning of the key variables (differential delay and criterion extents) is left rather vague when considering the SJ task from a neural perspective. Indeed, a more detailed neural-process model would be required in order to provide a concrete link between these variables and the operation of particular regions of the brain, such as the primary sensory cortices. A change in differential delay implies that the transformation between the objective SOA and the subjective SOA has changed. However, because we do not know exactly how the subjective SOA is represented in the brain (e.g. by the timing of the evoked activity, and/or by changes in the magnitude of a neural response, and/or by a population code in which different times activate different units, and/or by some other possibility) we cannot ascribe this variable exclusively to an early stage of processing. Similarly, because we do not know exactly how a criterion might be applied to a representation of the SOA (e.g. by thresholding applied to neural integration at one or several stages, or plastic changes in connections between temporal representations and other areas generating a response), we cannot claim unequivocally that this variable reflects late neural processing. However, despite these limitations the model does seem to represent a reasonable starting point from which to examine judgements of temporality.

---

[2] We thank an anonymous reviewer for this suggestion.

4.3. General implications

We have focussed our discussion on the phenomenon of temporal recalibration, because we have used that phenomenon to generate a data set that reflects an apparent temporal distortion. Hence our specific interpretation regarding shifts in a single criterion should not be generalised beyond this phenomenon. However, the interpretative issues we have identified are relevant to all studies that have inferred different neural processing times on the basis of discrepant PSS estimates obtained in different experimental conditions (e.g. Bartels & Zeki, 1998; Stone et al., 2002; Sugita & Suzuki, 2003). In short, we believe that such data do not provide unequivocal evidence for a change in neural processing times. Instead, it is equally plausible that such data are driven by changes in the criteria inherent in subjective timing judgments. Thus, in conclusion, we believe that fitting a model like the one used here could inform numerous debates concerning the relationship between timing perception and neural processing times. The fundamental message we would like to convey is that SJ tasks, just like TOJs, are not well suited for drawing strong conclusions about the mechanisms underlying apparent timing distortions. They encompass an ambiguity, reflecting a change in neural processing times, a change in one of two criteria used to differentiate synchrony from asynchrony, or some combination of these two possibilities.

Acknowledgments

D.H.A. is supported by an Australian Research Council discovery project grant and fellowship. K.Y. visited D.H.A.'s lab to discuss these ideas thanks to a Royal Society travel grant. N.J. collected these data whilst completing a Nuffield Foundation summer internship. The authors would like to thank John Burkardt for making his Matlab Nelder–Mead simplex code available online.

References

Allan, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception and Psychophysics, 18*, 29–36.

Allik, J., & Pulver, A. (1994). Timing of visual events for motion discrimination. *Vision Research, 34*, 1585–1594.

Baron, J. (1969). Temporal ROC curves and the psychological moment. *Psychological Science, 15*, 299–300.

Bartels, A., & Zeki, S. (1998). The theory of multistage integration in the visual brain. *Proceedings of the Royal Society of London Part B: Biological Sciences, 265*, 2327–2332.

Bedell, H. E., Chung, S. T., Ogmen, H., & Patel, S. S. (2003). Color and motion: Which is the tortoise and which is the hare? *Vision Research, 43*, 2403–2412.

Burr, D., Silva, O., Cicchini, G. M., Banks, M. S., & Morrone, M. C. (2009). Temporal mechanisms of multimodal binding. *Proceedings of the Royal Society of London Part B: Biological Sciences, 276*, 1761–1769.

Clifford, C. W., Arnold, D. H., & Pearson, J. (2003). A paradox of temporal perception revealed by a stimulus oscillating in colour and orientation. *Vision Research, 43*, 2245–2253.

Cravo, A. M., Claessens, P. M., & Baldo, M. V. (2011). The relation between action, predictability and temporal contiguity in temporal binding. *Acta Psychologica, 136*, 157–166.

de Lange, H. (1958). Research into the dynamic nature of the human fovea–cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America, 48*, 777–784.

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision, 9*, 7–16.

Eagleman, D. M., & Sejnowski, T. J. (2000). Latency difference versus postdiction: Response to Patel et al.. *Science, 290*, 1051a.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*, 773–778.

Gibbon, J., & Rutschmann, R. (1969). Temporal order judgement and reaction time. *Science, 165*, 413–415.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grove, P., Ashton, J., Kawachi, Y., & Sakurai, K. (2009). Sensory and decisional factors in the resolution of stream/bounce displays. *Journal of Vision, 9*(8), 723.

Hanson, J. V., Heron, J., & Whitaker, D. (2008). Recalibration of perceived time across sensory modalities. *Experimental Brain Research, 185*, 347–352.

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research, 186*, 517–524.

Heron, J., Roach, N. W., Whitaker, D., & Hanson, J. V. (2010). Attention modulates the plasticity of multisensory timing. *European Journal of Neuroscience, 31*, 1755–1762.

Kristofferson, A. B. (1967a). Attention and psychophysical time. *Acta Psychologica, 27*, 93–100.

Kristofferson, A. B. (1967b). Successiveness discrimination as a two-state, quantal process. *Science, 158*, 1337–1339.

McDonald, J. J., Teder-Salejarvi, W. A., Di Russo, F., & Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience, 8*, 1197–1202.

Miller, J., & Schwarz, W. (2006). Dissociations between reaction times and temporal order judgments: A diffusion model approach. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 394–412.

Mollon, J. D., & Perkins, A. J. (1996). Errors of judgement at Greenwich in 1796. *Nature, 380*, 101–102.

Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London Part B: Biological Sciences, 264*, 393–399.

Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London Part B: Biological Sciences, 264*, 1407–1414.

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences of the USA, 106*, 9169–9173.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*, 308–313.

O'Neill, R. (1971). Algorithm AS 47: Function minimization using a simplex procedure. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 20*, 338–345.

Rosenberger, W. F., & Grill, S. E. (1997). A sequential design for psychophysical experiments: An application to estimating timing of sensory events. *Statistics in Medicine, 16*, 2245–2260.

Roufs, J. A., & Blommaert, F. J. (1981). Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique. *Vision Research, 21*, 1203–1221.

Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive Psychology, 47*, 333–366.

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychological Science, 12*, 205–212.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General, 130*, 799–832.

Stelmach, L. B., & Herdman, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 539–550.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629–686). London: Academic Press.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al (2002). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London Series B: Biological Sciences, 268*, 31–38.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature, 421*, 911.

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception and Psychophysics, 42*, 224–239.

van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception and Psychophysics, 70*, 955–968.

Venables, P. H. (1960). Periodicity in reaction time. *British Journal of Psychology, 51*, 37–43.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *Journal of the Acoustical Society of America, 66*, 1364–1380.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research: Cognitive Brain Research, 22*, 32–35.

Weisenberger, J. M. (1986). Sensitivity to amplitude-modulated vibrotactile signals. *Journal of the Acoustical Society of America, 80*, 1707–1715.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics, 63*, 1293–1313.

Zampini, M., Shore, D. I., & Spence, C. (2005). Audiovisual prior entry. *Neuroscience Letters, 381*, 217–222.