Original Article

QJEP

Quarterly Journal of Experimental Psychology
2021, Vol. 74(2) 312–325

DOI: 10.1177/1747021820959756
qjep.sagepub.com
SAGE

Action at a distance: Long-distance rate adaptation in event perception

Christopher C Heffner[1,2,3,4], Rochelle S Newman[1,2] and William J Idsardi[1,3]

Abstract

Viewers' perception of actions is coloured by the context in which those actions are found. An action that seems uncomfortably sudden in one context might seem expeditious in another. In this study, we examined the influence of one type of context: the rate at which an action is being performed. Based on parallel findings in other modalities, we anticipated that viewers would adapt to the rate at which actions were displayed at. Viewers watched a series of actions performed on a touchscreen that could end in actions that were ambiguous to their number (e.g., two separate "tap" actions versus a single "double tap" action) or identity (e.g., a "swipe" action versus a slower "drag"). In Experiment 1, the rate of actions themselves was manipulated; participants used the rate of the actions to distinguish between two similar, related actions. In Experiment 2, the rate of the actions that preceded the ambiguous one was sped up or slowed down. In line with our hypotheses, viewers perceived the identity of those final actions with reference to the rate of the preceding actions. This was true even in Experiment 3, when the action immediately before the ambiguous one was left unmodified. Ambiguous actions embedded in a fast context were seen as relatively long, while ambiguous actions embedded in a slow context were seen as relatively short. This shows that viewers adapt to the rate of actions when perceiving visual events.

Keywords

Event perception; visual perception; rate; time perception

Received: 22 October 2019; revised: 11 August 2020; accepted: 17 August 2020

Imagine being in the spectators' gallery as an important piece of legislation is considered. During an intense debate, one member presses an elbow to the side of the legislator next to her. Was that a playful act of *nudging*, showing a bond between the two, or an aggressive *elbowing*, indicative of tension? Since nudging appears slower than elbowing, the two actions might be distinguished based, in part, on the speed that the elbow is being flung. In this article, we examine the influence of action rate on the perception of visual actions. We focus on rate *adaptation*, the idea that the rate of an action is perceived relative to its context.

Actions can be distinguished from one another using a variety of different cues. Both bottom-up sensory and perceptual information and top-down knowledge of the world combine to allow viewers to interpret the actions of others (Blake & Shiffrar, 2007; Zacks, 2004). Duration and speed can help distinguish actions. Like nudging and elbowing, other actions can be distinguished by the speed at which an action takes place, as well as how long it takes an action to be performed. For example, when viewers were asked to segment the random motions of two objects into distinct actions, motion cues contributed significant variance to the segmentation points perceived in the scene (Zacks, 2004), with these points accompanied by neural activation in the MT complex, also known as V5 (Zacks et al., 2006).

The perception of time for actions is remarkably context dependent. Viewers judge the duration of a visual stimulus on the screen as shorter when a nearby object flickers quickly than when it flickers slowly (Johnston et al., 2006). Speed interacts with perceived duration,

[1]Program in Neuroscience & Cognitive Science, University of Maryland, College Park, MD, USA
[2]Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, USA
[3]Department of Linguistics, University of Maryland, College Park, MD, USA
[4]Department of Communicative Disorders and Sciences, University at Buffalo, Buffalo, NY, USA

Corresponding author:
Christopher C Heffner, Department of Communicative Disorders and Sciences, University at Buffalo, 122 Cary Hall, Buffalo, NY 14215, USA.
Email: ccheffne@buffalo.edu

although whether faster actions are perceived as being relatively long or short depends on the task and the expertise of the viewers (Kaneko & Murakami, 2009; Sgouramani & Vatakis, 2014). Actions are also perceived as relatively long when they are seen as being made up of a large number of separate events rather than a small number (Faber & Gennari, 2015; Liverence & Scholl, 2012).

One type of context that viewers seem particularly sensitive to is whether actions are biologically grounded. The perceived duration of two static images of bodies in motion respects the time it would take for the body depicted to link the poses shown in each static image, meaning that the perception of timing is dependent on biological realities of motion (Orgs & Haggard, 2011; Shiffrar & Freyd, 1993). Viewers can more accurately estimate the endpoint of an action when that action is presented as originally recorded than when the velocity of the action is sped up or slowed down, especially when the actions are perceived as being biologically generated (Martel et al., 2011). Indeed, it has been proposed that viewers have separate timekeepers for biological actions compared to non-biological actions (Carrozzo & Lacquaniti, 2013; Carrozzo et al., 2010; Pozzo et al., 2006). This highlights the durational component of the action perception system as being one that is acutely aware of the context of actions being performed.

In this study, we test the hypothesis that viewers may adapt to the timing properties of early-occurring actions in a sequence to influence their interpretation of the temporal properties of later-occurring actions in a sequence. We focus on the rate of actions, examining whether viewers' experience of the speed of a set of actions can bleed into the interpretation of later-occurring actions. If viewers are indeed using context to inform the *content* of later-occurring visual information in a signal, they may also use it to infer aspects of the *duration* of later-occurring content. Thus, changing the rate of a signal may influence the perceived content of a later-occurring signal. In the event perception literature, this is best supported by the literature concerning predictive processing in events. However, it can also be expected through mechanisms of contrast or even backward inference.

It has been suggested that context informs the perception of timing in actions in part because viewers make predictions about future actions when perceiving a signal. Under these proposals, viewers are actively predicting the outcomes or goals of actions, ensuring accurate perception by projecting actions forward in time (Schütz-Bosbach & Prinz, 2007). For timing, this means that viewers are forming expectations about the length and form of actions. Viewers are better able to predict actions when they are congruent with previously established timing regularities. In one task, participants viewed a video clip of a point-light action, where the actions of real-life actors in a space too dark to see them were depicted using points of light at key locations on the body. At one point, the video cut out and was replaced with a blank screen, which had one of three possible durations. After the blank screen disappeared, the action either continued in a way that was congruent with previous footage, showing the same action, or incongruent, showing a different action. Viewers were better at telling whether an action was congruent when the action picked up at a point corresponding to the duration of the blank screen rather than when the subsequent scene started earlier or later. That is, viewers are forecasting the temporal dynamics of the location of the actors in the video clip and using that to more effectively judge congruence (Graf et al., 2007). This is also true for more naturalistic videos (Verfaillie & Daems, 2002). Participants seem to be predicting the goal of point-light actions (Elsner et al., 2012) and can use these predictions to better overcome visual noise than if they were unable to make such predictions (Parkinson et al., 2011).

The idea that prediction is key for action perception is perhaps best exemplified in investigations of representational momentum. Representational momentum is the idea that viewers' memory of actions seems to be biased in favour of the direction that a scene was unfolding before the time of test. The recollection of the orientation of a rotating rectangle, for example, seems to be biased in the direction of rotation (Freyd & Finke, 1984). Similar patterns have been observed for dot patterns (Finke & Freyd, 1985), simple geometric patterns (Bertamini, 1993), and a simulated visual landscape similar to that perceived out of a car on a highway (Thornton & Hayes, 2004). This is not just limited to the direction that objects are moving in; viewers are also more likely to recall actions in directions that are biased towards the goals of the actors in a visual scene even above and beyond simple perceptual momentum (Hudson et al., 2016).

These predictive mechanisms may be supported using action production mechanisms in perception. Much of this speculation centres on the use of mirror neurons, which describe a subset of neurons that are said to fire both when an action is performed and when it is observed or imagined (Rizzolatti & Craighero, 2004). Although not uniformly accepted (Decety & Grèzes, 1999; Pavlova et al., 2003), it has been hypothesised that action production mechanisms may be involved in simulating the outcome of future actions (Prinz, 1997). These connections include aspects of timing. Viewers are more accurate at perceiving the difference between fast and slow gaits when those gaits are physically plausible than when they are physically impossible, suggesting that viewers' experience of that motion is dependent on their knowledge of the physical constraints of walking (Jacobs et al., 2004). The interaction between timing and production seems to work in both directions; preparing an action seems to subjectively slow the passage of time, with the duration of simple visual displays being perceived to be relatively long when an unrelated motor action is in the process of being prepared (Hagura et al., 2012).

However, prediction-based accounts are not the only ones that suggest that viewers may respond to temporal context when interpreting actions. Studies of temporal aftereffects indicate that the durations of simple sensory events, such as illuminated lights, are perceived relative to the temporal context in which they are found. The duration of a fixed reference light is perceived to be long relative to its true duration when it is embedded in a sequence of short lights; the duration of the same reference light embedded in a sequence of long lights is perceived to be relatively short (Walker et al., 1981). They suggested that these aftereffects are the result of adaptation on the part of duration-based feature detectors. Repeated presentation of stimuli occurring at a single duration would lead to the exhaustion of the neural population dedicated to detecting that duration, causing the neural populations picking up on later, similar durations to be skewed in the opposite direction. These hypotheses were later formalised into modelling work suggesting that the channels occur relatively early in processing at a modality-specific level (Heron et al., 2012). Later testing of these durational aftereffects suggests that these results can transfer across visual hemifields when occurring due to the repeated presentation of a stimulus at a single duration (Li et al., 2015), although not with a solitary comparison stimulus (Johnston et al., 2006; Ortega et al., 2012; Zhou et al., 2014). Under these accounts, viewers' perception of final actions within a sequence might be affected by simple sensory aftereffects.

Perception can also engage higher level cognition without necessarily being predictive; backward inference could also be used to support the idea that the rate of early-occurring information can influence perception of later content in a signal. Although this idea is relatively new to the world of event perception (Papenmeier et al., 2019), it has been well-established in the reading literature (Graesser et al., 1994). The idea is that viewers make inferences about the behaviour they saw earlier in a clip based on later-occurring information. Papenmeier et al. (2019) presented viewers with clips of a soccer match where an anticipated and crucial point of contact between a ball and a player was either present in the clip or not present. The clips could either be followed with video information that was in line with the idea that the contact occurred, that was not in line with that idea, or with a blank mask. When subsequent context was present, viewers largely behaved in line with it, falsely recalling ball contact when the subsequent context indicated that it was present. However, in conditions where the mask was present, viewers did not give responses in line with the content of the previous context. It seems that they were relying solely on subsequent context, rather than predictive processing, to infer the presence of the missed ball contact. Likewise, when viewing simple three-panel comics, viewers showed evidence that they were inferring the content of missing initial and medial panels, but not missing final panels, indicating, again, that later information was used to fill in previous, missing content (Magliano et al., 2017). Under backward inference accounts, viewers may not be making predictions about the rate of subsequent actions, but instead making inferences based on later-occurring information.

Within vision, the temporal recalibration literature provides support for the idea that the interpretation of temporal information can depend on the immediate context. In a typical audio-visual recalibration task, participants are given an exposure period with a combination of two signals: one visual and one auditory. The signals are presented with a uniform time lag between them. For instance, the visual cue always precedes the audio one, or vice-versa. When tested later in the experiment, participants are asked to evaluate the lag between test visual and auditory stimuli. The perception of the later-occurring information is skewed in the direction of the training information; items with a lag resembling the earlier information are perceived as occurring at identical time points (Vroomen et al., 2004). This suggests that viewers are recalibrating their perception of temporal synchrony based on their prior temporal context. Similar results have been shown for tactile-visual asynchrony (Keetels & Vroomen, 2008) and even for asynchronies reflecting multiple aspects of a single visual stimulus, including changes in colour and in direction of motion (Arnold & Yarrow, 2011). More recently, temporal audio-visual recalibration effects have been shown on a more rapid scale. Some studies have shown effects of a single, previous trial on perceived simultaneity (Noel et al., 2016; Van der Burg et al., 2013). Rapid recalibration adds to, rather than contrasts with, longer term recalibration effects (Van der Burg et al., 2015), and can be indexed by event-related potential (ERP) responses (Simon et al., 2017). Thus, listeners adjust their temporal responses based on even the information gained from a single trial of exposure to interpret ambiguous audio-visual experiences, suggesting that timing information can quickly affect the perception of visual events.

The idea that timing information at the beginning of a sequence should influence the perception of actions at the end of one is additionally supported by studies in the auditory domain. Earlier-occurring timing information can change the perception of a ball rolling down a steep ramp into a ball rolling down a shallow ramp (Fowler, 1990), make a [w] sound be perceived as a [b] sound (Miller & Liberman, 1979), or make the phrase "minor or child" sound like "minor child" (Dilley & Pitt, 2010). In the auditory domain, this phenomenon is generally known as "rate adaptation." Roughly analogous experiments in the visual domain have shown that changes in velocity have important effects on representational momentum (Martel et al., 2011). The perceived speed of actions, as modulated by the constraints of biological motion, can often change the perceived duration of simple visual frames surrounding the visually presented actions (Orgs et al., 2011, 2013).

All told, this evidence suggests that viewers may use temporal regularities in the context of a signal to determine what actions are being performed. However, it is not yet clear how relative contrasts in duration or perceived rate might lead to changes in how ambiguous actions might be perceived. In the present experiments, we created action sequences from clips of an actor interacting with a touchscreen. We focus on actions that were ambiguous in how they were labelled by viewers ("identification") or in whether they were segmented into two separate actions or perceived as a single one ("segmentation"). Key sequences ended in critical actions that were temporally ambiguous either to their segmentation or to their identity.

First, in Experiment 1, we tested whether the critical actions that we used were truly ambiguous: for example, that a sequence of two "tap" actions could also be segmented as a single "double tap" action based on the rate of the individual taps on the screen, or whether an ambiguous action could be identified variously as a "swipe" or a "drag" action depending on its duration. The rate information of the context before the critical actions was then modified to assess the effects of the preceding action rate on the perception of the critical actions. In Experiment 2, the rate of all but the critical actions was modified; in Experiment 3, the rate of only the context far away from (nonadjacent to) the signal was manipulated to examine whether any context effects observed in Experiment 2 were the result of only the immediate context, or whether they could arise from contrasts with rate information that would be further removed. This would push the existence of rate effects into the temporal domain of "cognitively-mediated" temporal perception of visual information, further than 500 ms or so (Rammsayer, 1999). If rate adaptation effects in speech (e.g., Dilley & Pitt, 2010) are any guide, viewers should perceive the critical actions relative to the rate of the context actions in both cases.

Experiment 1

Before we could see whether these touchpad actions could be affected by the rate of the context actions around them, it was first necessary to see whether they were rate-dependent at all. Experiment 1 was thus designed to test whether the rate at which touchscreen actions are performed can affect which actions are perceived.

Method

Participants. A total of 40 participants completed the experiment. Since the effect size for this study and subsequent ones was not available before this one was run, the sample size for this study was chosen based on analogous studies in the speech perception literature, as well as sample sizes for other tasks of action perception. Four of those participants were excluded: two for at-chance performance across trials (indicating a failure to attend to or understand instructions) and two for missing demographic information. This left 36 native-speaking participants (15 females, 20 males, 1 other) at least 18 years of age (M=21.3, range=18–31) and with no history of uncorrected vision impairments. Participants, recruited from the University of Maryland (UMD), College Park community, were compensated with US$5 for the 30- to 45-min experiment. This and all other studies in this article were approved by the UMD Institutional Review Board (IRB).

Materials. Participants saw 42 experimental videos and 84 filler videos. The videos were recorded using a fixed digital camera at 24 fps showing a single, seated actor interacting with a touchscreen device, with a camera angle chosen such that the movement of the actor's fingers on the touchscreen was the primary cue available to the actions being performed. Each video included a sequence of either seven or eight actions; possible actions included a tap, a press, a drag, a swipe, a double tap, a twist or rotate,[1] a pinch, or a spread. Individual actions generally were 1s long, with some variation depending on the action and the individual production. The actor was instructed to produce the actions as a fluid sequence of independent actions; she was not instructed to return to centre or pause between actions.

All experimental items ended in one of two possible action sequences: a drag action or two tap actions.[2] The total length of this critical region before modifications was between 0.54 and 1.33 s (M=0.93 s), preceded by a precursor region of between 5.38 and 7.75 s (M=6.51 s). The drag or (two) tap actions could be ambiguous in their timing properties with a swipe action and a double tap action, respectively.[3] All experimental items were rate-modified using a free software package, ffmpeg (https://www.ffmpeg.org/). Critical actions were sped up and made ambiguous by dropping two of every three frames. Filler items ended with one of the other actions—twist, spread, pinch, swipe, press, or double tap—and were not rate-modified. Participants generally perceived the filler actions accurately (M=75.1%, with chance at 12.5%).

The manipulations for this experiment are illustrated in Figure 1. The original actions were recorded at a relatively slow rate, one that was not ambiguous as to action identity. To see if the rate of the sequence could impact perception, the final action in the series was sped up at one of three different rates: 66% of the original duration (a rate 150% the speed of the original), 50% of the original duration (a rate 200% the original speed), or 33% of the original duration (a rate 300% the original). Items could not be sped up further without visual artefacts. If rate can impact perception, we would expect that some of these sped up versions would be identified as a different action (e.g., a swipe rather than a drag).

Procedure. The present experiment used a 3 (critical rate: 150%, 200%, 300%) $\times$ 2 (type: tap and drag) design for the experimental items, with participants sorted into three

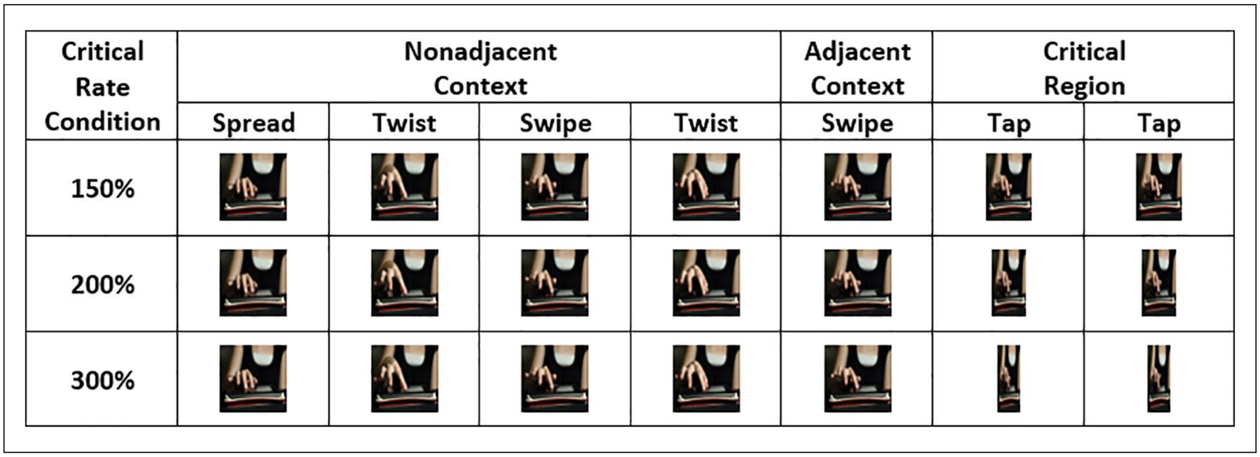

Figure 1. A schematic depiction of an action sequence that shows the rate manipulation of Experiment 1. Only the rate of the critical region is modified.

PRESS 1	TAP 2	DRAG 3	SWIPE 4
DOUBLE TAP 5	ROTATE 6	PINCH 7	SPREAD 8

Figure 2. The response mapping for the study as shown to participants on each trial.

equally sized counterbalanced groups that ensured that each item showed up across the three lists in each critical rate. Each participant thus saw seven items with each combination of critical rate and item type. Participants were told to indicate which action they saw immediately prior to a sine-wave tone (described as a "beep") that was inserted within the action sequence. They were told the beep would usually occur at the end of the trial but could occur anywhere in the sequence of actions; they were not told anything about the length of each sequence. For filler trials, the tone could occur anywhere between the end of the second action and the final action, with more tones occurring towards the end of the sequence. This was done to encourage participants to attend to the entire sequence on all trials, as they could not be sure before any single trial started whether they would need to wait to the end of the sequence. For all experimental trials, the tone was placed at the end of the clip. Participants were told to indicate which action they saw immediately prior to a tone by pressing a button between 1 through 8 on a keyboard, with possible responses listed on the screen and matched to a number, as shown in Figure 2. Trial order was randomised by participant. After initial presentation, participants could repeat each trial up to twice before responding, leading to a total of three presentations.

Analysis. First, inaccurate trials were discarded. Tap trials were removed if the event reported was not a tap or a double tap (9% of trials) and drag trials were removed if the event reported was not a drag or a swipe (7%). For accurate trials, responses were coded as a "long response" if they were reported as originally recorded, and as a "short response" if the short analogue of each original event was reported (i.e., double tap for tap trials or swipe for drag trials). For analysis of the critical rate manipulation, the rates were expressed in relation to the original rate (150%, 200%, or 300%). Coding this factor in a continuous fashion allowed for relationships between each factor level to be considered in the model. Mixed models implemented in the lme4 package (Bates et al., 2016) and refined using the RePsychLing package (Baayen et al., 2015) within R (version 3.3.1) were used to analyse the dataset. To aid in model convergence, the BOBYQA algorithm was used to implement the mixed models. Effect size and power estimates are based on observed variance in random slopes and factor levels (Brysbaert & Stevens, 2018).

Results

Figure 3 shows the extent to which critical rate effects the perception of the critical region. The fastest rate is the least

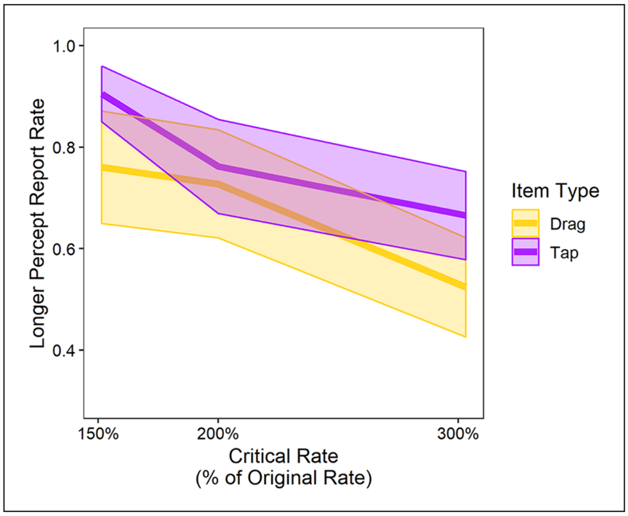

Figure 3. Experiment 1 results. The semi-transparent ribbon shows 95% confidence intervals on a by-participant basis.

likely to lead viewers to see the event as long, while the fast rate shows the highest likelihood of being perceived as long. When the critical action seems relatively fast compared to the context, it is less likely to seen as a long action.

To evaluate these findings, an initial model was developed to determine the importance of various random effects in this model. The initial model included random intercepts by participant and by item, random slopes for item type by participant, and random slopes for critical rate by item and by participant. These random slopes by item were particularly important because these materials have not been used before, and it is reasonable to assume that the two item types may have different baselines both in terms of how often they are seen as one action or the other and in terms of the effect sizes of the rate manipulations. The RePsychLing package (Baayen et al., 2015) was used to determine the maximum number of supported random effects structure. A principal component analysis (PCA) was performed on the variance-covariance matrix of the initial model. The resulting PCA indicated that a single component was supported by item and two components were supported by participant for this dataset. Despite this, dropping the random slope for critical rate by item did decrease model fit, $\chi^2(1)=11.3$, $p<.001$, as such, it was included in the intermediate model. Cutting the random slope for critical rate by participant, however, did not decrease model fit, $\chi^2(1)=0$, $p=1$. Thus, the intermediate model included the random slopes for item type and critical rate by participant as well as random intercepts by item and by participant.

Model comparison was performed to determine the influence of fixed effects on model fit. Removing critical rate from the model (and its interaction with item type) significantly decreased model fit, $\chi^2(2)=72.7$, $p<.001$, suggesting that critical rate had an influence on the perception of the critical events. A relatively slow critical action led to the perception of a longer event. Removing item type from the model, however, did not decrease model fit, $\chi^2(2)=4.50$, $p=.11$, indicating that the item types did not vary in their baseline propensity to be seen as long or short; the tap action was equally likely to be seen as two separate taps as the drag action was to be seen as a drag. The resulting final model included effects of critical rate but not of item type (Table 1).

Table 1. The best-fitting model for Experiment 1.

Factor(s)	Estimate (*b*)	95% CI	*z*	*p*	*d*
Intercept	4.07	[3.31, 4.92]	10.3	<.001	–
Critical rate	−1.12	[−1.39, −0.862]	−8.57	<.001	−.39

Experiment 2

Experiment 1 indicated that rate is a factor in guiding the perception of these touchscreen actions. With that in mind, it is now possible to examine the extent to which the perception of the rate of an ambiguous action can be influenced by its context. Experiment 2 was designed to examine whether context rate adaptation is possible in event perception. As such, "context" was defined in as permissive of a sense as possible, with the rate of all actions other than the critical ones being modified.

Method

Participants. A total of 41 participants (15 females, 26 males), aged 18–26 years ($M=20.4$) successfully completed the experiment. Participants, recruited from the UMD community, were compensated with US$5. The experiment usually lasted around 30 min, although some participants took up to 45 min.

Materials. Participants saw 63 experimental videos and 63 filler videos. Experimental items played ended in one of three possible action sequences: a drag action, two tap actions, or a press action. The unmodified versions of all but the press actions were identical to the videos used in Experiment 1. The press actions were used as fillers in Experiment 1, with a tone placed somewhere in the middle of the clip; for Experiment 2, the tone was played at the end. The rate of the actions in the critical region were sped up to 300% of the original speed, the fastest duration in Experiment 1. This led to a range of critical region durations from 0.17 to 0.43 s ($M=0.29$ s). This was the duration that led to the most ambiguous possible perception. Filler items ended with one of the other actions: twist, spread, pinch, swipe, press, or double tap. Filler actions were generally perceived accurately ($M=83.1\%$, chance at 12.5%).

Precursor actions (e.g., all actions prior to the critical ones) within each experimental item were also modified. For some items, the duration was kept unmodified; when

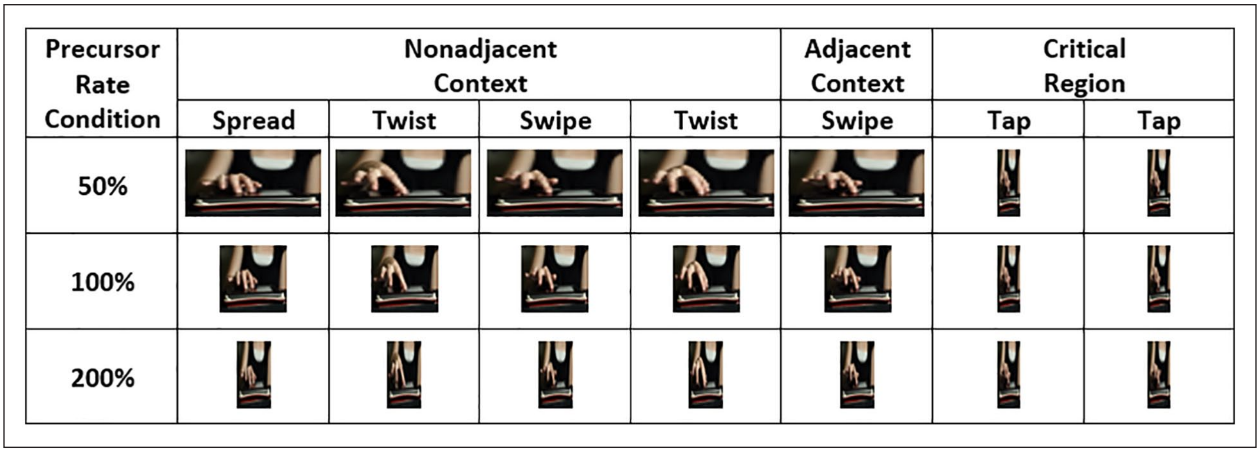

Figure 4. A schematic depiction of an action sequence that shows the rate manipulation of Experiment 2. The rate of the critical region is speeded to make it ambiguous, while the rate of both the adjacent and nonadjacent context are modified to examine the influence of precursor rate on event perception.

unmodified, these items had precursor durations ranging between 5.38 and 7.83 s (M=6.42 s). For others, the rate was either doubled (by dropping every other frame) or halved (by duplicating every frame), as depicted in Figure 4, below. Thus, there were three levels for the precursor rate, expressed as a percentage of the rate in the original recordings: 50% (slowed), 100% (unmodified), and 200% (speeded). To keep the experimental items from standing out, filler items were presented with uniform but analogous durational properties, with rates kept unmodified, doubled, or halved.

Procedure. The experiment used a 3 (precursor rate: 50%, 100%, and 200%) $\times$ 3 (type: tap, drag, and press) design for the experimental items, with three lists that counterbalanced items to precursor rates. The 63 experimental items were equally distributed across each type of precursor rate and item type, meaning that there were 7 items per combination of precursor rate and item type. The procedure was otherwise identical to Experiment 1.

Analysis. First, inaccurate trials were excluded. This led to the removal of 8% of tap trials and 7% of drag and press trials. Responses were coded identically to Experiment 1. The precursor rate was coded continuously in relation to the original rate. Mixed models were employed, using a combination of the lme4 package (Bates et al., 2016) and the RePsychLing package (Baayen et al., 2015) within R (version 3.3.1).

Results

Figure 5 shows that viewers adapted to the rate of context actions to determine what they saw in ambiguous actions regardless of whether the "press" actions are included. A principal components analysis (PCA) was performed on the variance–covariance matrix of the fully specified model in the RePsychLing package (Baayen et al., 2015).

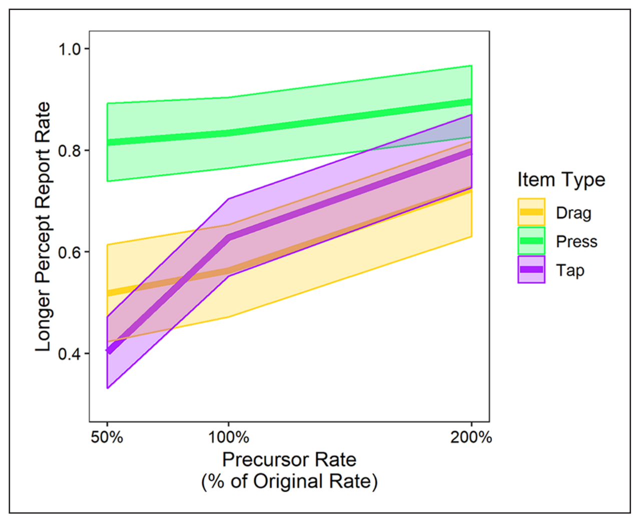

Figure 5. Experiment 2 results, showing the effects of precursor rate and item type on the likelihood that participants gave a long response. The proportion of long responses was rate-dependent, particularly for the "tap" action; people were more likely to perceive the actions as short when the context was relatively long (i.e., slow). The semi-transparent ribbon shows standard errors on a by-participant basis.

This PCA indicated that all the random components for participant and all but one of the components for item type in the model were plausible to include. Indeed, removing the random slope for precursor rate by item did not decrease model fit, $\chi^2(1)=0.0398$, $p=.84$, indicating that it could be safely removed from the model. The full model without the random slope for precursor rate by item will be referred to as the intermediate model.

The effects of precursor rate and item type were examined by comparing the intermediate model to models that lacked both the main effects of each factor and the interaction between them. Removing the effects of precursor rate,

Table 2. The best-fitting model with the press actions included.

Factor(s)	Estimate (*b*)	95% CI	*z*	*p*	*d*
Intercept	−0.525	[−1.33, 0.259]	−1.33	.18	–
Precursor rate	0.924	[0.592, 1.28]	5.50	<.001	.31
Type: press	1.93	[1.00, 2.88]	4.12	<.001	.65
Type: tap	−0.635	[−1.60, 0.334]	−1.31	.19	.21
Precursor rate × type: press	−0.316	[−0.801, 0.178]	−1.32	.19	.11
Precursor rate × type: tap	0.692	[0.225, 1.15]	3.10	.002	.23

Drag actions were used as the baseline.

Table 3. The best-fitting model in Experiment 2.

Factor(s)	Estimate (*b*)	95% CI	*z*	*p*	*d*
Intercept	−0.543	[−1.37, 0.192]	−1.33	.18	–
Precursor rate	0.945	[0.609, 1.31]	5.52	<.001	.34
Type: tap	−0.546	[−1.50, 0.421]	−1.14	.26	.20
Precursor rate × type: tap	0.602	[0.158, 1.05]	2.72	.007	.22

Drag actions were used as the baseline.

$\chi^2(3)=73.4$, $p<.001$, item type, $\chi^2(4)=43.3$, $p<.001$, and the interaction between them, $\chi^2(2)=18.2$, $p<.001$, all decreased model fit. Slowing the rate of a precursor action sequence leads to the perception of a final action as relatively fast; speeding the rate of a precursor action sequence leads to the perception of a final action as relatively slow. There was a significant difference in the baseline propensity of each type of action to be seen as long. Furthermore, the two effects interacted with each other, such that some actions had stronger precursor rate effects than others. The intercepts in the best-fitting model, shown in Table 2, indicate a small effect of precursor rate on the perception of the actions. The strongest deviation from the drag actions are the tap actions, which have stronger precursor rate effects.

Performing the same analysis without the press actions, to maintain continuity with the other experiments in this article, yielded similar results. Following the approach to random factors outlined above, the same random effects structure fit best for this analysis, yielding an intermediate model with all random intercepts and slopes other than the random slope for precursor rate by item, a model referred to as the intermediate model. The comparison between the intermediate model and the models lacking each individual effect was significant for the model lacking effects of precursor rate, $\chi^2(2)=69.1$, $p<.001$, and item type, $\chi^2(2)=7.14$, $p=.03$. Comparison of the intermediate model to one lacking the interaction between the two factors suggests that the two factors interacted with each other, as taking out the interaction decreased model fit, $\chi^2(1)=7.10$, $p=.008$; Table 3 shows the fixed model parameters for the best-fitting model. The computed d value for precursor rate suggests that this is a small effect size.

Experiment 3

Experiment 2 demonstrated evidence for rate adaptation in event perception. In Experiment 3, we set out to find whether the effects observed in Experiment 2 arose solely from an immediate contrast or whether earlier actions in the sequence had an effect. This is analogous to the literature from speech perception, where the nonadjacent context is labelled the "distal context" (Dilley & McAuley, 2008). Studying the nonadjacent context alone will allow us to see the time course of these adaptation effects: whether viewers aggregate information from a long timescale around an ambiguous action, or whether information is only retrieved and used within a short time window.

Method

Participants. A total of 41 native English-speaking participants (13 females, 27 males, 1 other), aged 18–27 years ($M=20.2$) were compensated with US$5 for completing the experiment.

Materials. As before, there were 42 experimental ("tap" or "drag") stimuli, mixed with 82 filler stimuli. The unmodified versions of these stimuli were identical to Experiment 1. The duration of the action immediately preceding the critical action was left unmodified, with double taps treated as a single action. Unlike in Experiment 2, only the rate of the earlier, nonadjacent actions in the action sequence was changed, as depicted in Figure 6, below. The nonadjacent portion of the precursor region had an average duration between 4.42 and 6.88 s ($M=5.54$ s), while the adjacent portion had an average duration of between 0.71 and 1.33 s ($M=0.96$ s). Nonadjacent rates of 50% (slowed), 100%

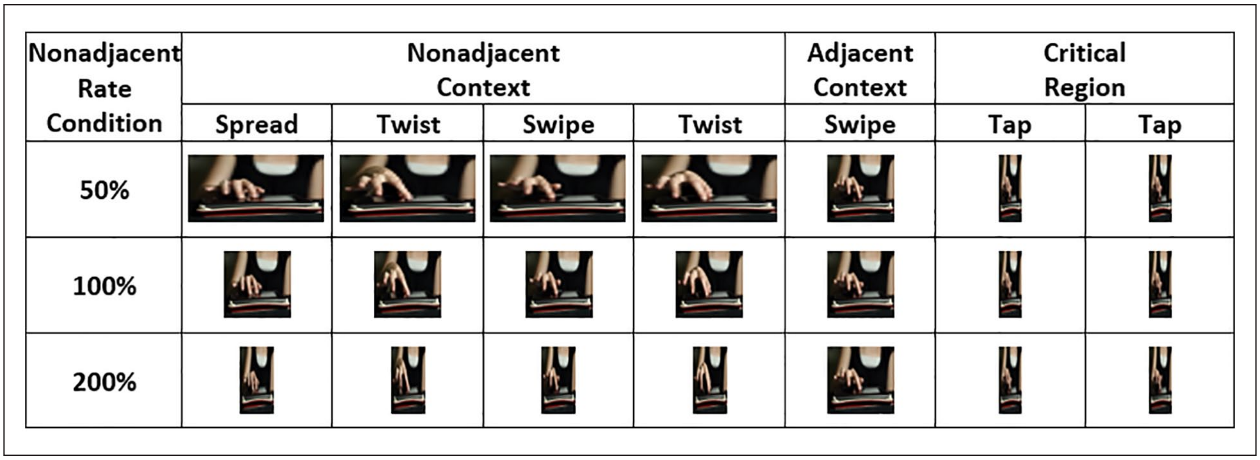

Figure 6. A schematic depiction of an action sequence that shows the rate manipulation of Experiment 3. The rate of the critical region is speeded to make it ambiguous, while the rate of just the nonadjacent context is modified to examine the influence of it alone on event perception.

(unmodified), and 200% (speeded) of original rate were used. These rates were identical to Experiment 2. Filler stimuli were generally perceived accurately (M = 74.5%, chance at 12.5%).

Procedure. The procedure was identical to Experiment 2 and Experiment 1, with items consisting of equal combinations of item type and nonadjacent context rate and lists counterbalancing the assignment of items to context rates.

Analysis. The analysis was identical to that in Experiments 1 and 2. Nine percent of responses for tap items and 5% for drag items were removed.

Results

The results for Experiment 3 are shown above in Figure 7. As before, a full model was developed for model comparison. A PCA indicated that at most one component by item and two by participant were sustainable according to the variation in the model. There was no significant difference between the full model and a model lacking the random slope of nonadjacent rate by item, $\chi^2(1) = 0$, $p = 1$, or between the full model and one without the random slope of nonadjacent rate by participant, $\chi^2(1) = 0$, $p = 1$. As such, the point of comparison to determine the influence of fixed effects was a model that includes random intercepts by item and by participant as well as random slopes for item type by participant.

First, the intermediate model was compared to models that lacked fixed effects of nonadjacent rate and item type to determine the influenced of those fixed factors on model fit. There was a significant decrease in model fit between the intermediate model and the one lacking fixed effects of nonadjacent rate (and its interaction with item type), $\chi^2(2) = 32.7$, $p < .001$. This confirms the pattern evident in

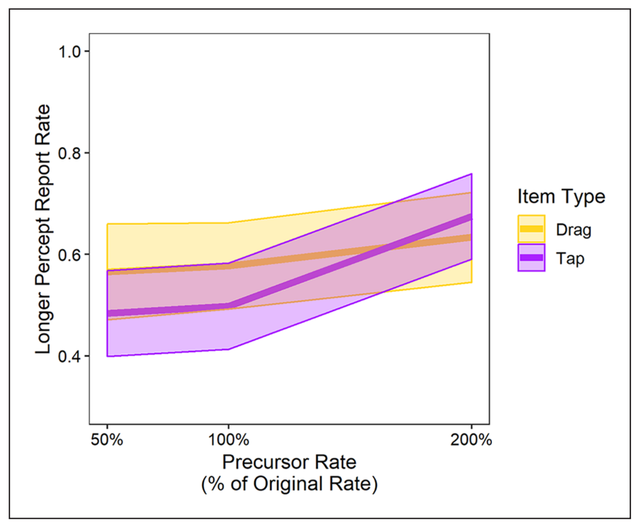

Figure 7. Experiment 3 results, showing the influence of nonadjacent rate on how often participants make a long response. Participants were influenced by the rate of the nonadjacent context, with relatively fast context rates leading to higher reports of seeing a longer percept, and relatively slow ones leading to lower reports. The semi-transparent ribbon shows standard errors on a by-participant basis.

Figure 7; event perception, like word segmentation, can be influenced by information at a distance. But there was no such decrease when comparing the intermediate model to the one lacking fixed effects of item type, $\chi^2(2) = 5.07$, $p = .08$, suggesting this was not limited to a particular action. The best-fitting model, with just a fixed effect of nonadjacent rate, is shown below in Table 4. Although the effects of nonadjacent rate are significant, they reflect a small effect size. To give a sense of the statistical power of this and the other experiments, we performed a subsequent power sensitivity analysis using the simr package in R

Table 4. The best-fitting model for Experiment 3.

Factor(s)	Estimate (*b*)	95% CI	*z*	*p*	*d*
Intercept	−0.223	[−0.763, 0.314]	−0.822	.41	–
Nonadjacent rate	0.527	[0.331, 0.728]	5.29	<.001	.20

(Green et al., 2016). The final model below was compared to one without an effect of nonadjacent rate (i.e., with just an intercept) at iteratively smaller effect sizes. The analysis indicated that effect sizes as small as 0.10 could have maintained a power of $\beta = 0.80$ using this combination of sample size and item numbers.

General discussion

We sought evidence for rate adaptation in event perception. To do this, sequences of touchscreen actions were created. Key sequences contained a critical ambiguity at the end of the sequence. Experiment 1 showed that perception of this ambiguity was, in part, dependent on the displayed speed of those actions. In Experiment 2, we manipulated the rate of the context actions to assess to what extent event perception was influenced by the relative difference between the rate of an ambiguous action and its preceding context. Viewers adapted to the rate of the context; when the context rate was slow in comparison to the rate of the critical action, viewers were more likely to perceive the critical action as relatively short. Conversely, when the context rate was faster than the rate of the critical action, viewers were more likely to perceive the critical action as relatively long. This was true even in Experiment 3, when the action immediately adjacent to the final ambiguous one had a duration that was held constant, showing that these effects hold even for longer timescales that are sometimes considered cognitively mediated, as opposed to shorter, sensory-based durations (Rammsayer, 1999).

To be sure, the number of actions examined in this study is small. We used only a few actions taken from a touchscreen; of the three chosen that were believed to be ambiguous, only two of them showed a substantial amount of ambiguity and were thus used in other testing. The pool of possibilities is limited by the low number of actions, performed using touchscreens or otherwise, that contrast solely based on duration. Subsequent studies will need to examine actions other than ones taken on a touchpad. However, this study provides an interesting first step in the direction of investigating the influence of rate adaptation on event perception, particularly given that the two actions that we used differed from one another in the type of ambiguity being considered. The tap action was ambiguous in terms of its segmentation, as the two taps that ended the sequence could also be seen as a single double tap event. The swipe action was ambiguous in terms of its identity, as it could also be seen as a drag. Therefore, although the number of actions was not high, the two actions represented different types of ambiguity in event processing, both of which appear to be affected by rate context. This reaffirms the idea that influential context can take a variety of forms. It has been abundantly demonstrated that aspects of context that rely on top-down knowledge, such as whether an action is biologically grounded (Orgs & Haggard, 2011; Shiffrar & Freyd, 1993), influence of those actions. But context can also be defined in terms of bottom-up information including the duration of actions themselves. This has been shown in the context of audio-visual recalibration (Van der Burg et al., 2013). As the recalibration literature and the present experiment reveal, bottom-up aspects of temporal context can quickly and strongly lead to a re-evaluation of subsequent material.

These results were generally not obtained for the "press" actions used in Experiment 2. We created those believing that they would be ambiguous. However, they were generally less ambiguous than the others; they were generally perceived as "press" actions rather than "tap" actions at a rate of at least 80% regardless of the context rate. Indeed, even in filler actions, "press" and "tap" actions were rarely confused with each other, while, say, "drag" actions were seen as "swipe" actions about 30% of the time. It is likely that viewers were picking up on other cues to action identity besides the rate that the actions were performed at, which led the stimuli to be seen as unambiguous. For that reason, they were not used as experimental stimuli in Experiments 1 and 3.

Our findings provide expanded evidence for the types of context that can inform viewers' perception of visual actions, which, in turn, has implications for theories of how that context is integrated. These results are concordant with the idea of active prediction on the part of viewers when watching actions (Schütz-Bosbach & Prinz, 2007). Under a prediction-based account of these results, viewers are not only making predictions about how long actions should take (Elsner et al., 2012; Graf et al., 2007; Parkinson et al., 2011); they are also using those predictions to disambiguate later-occurring information in the signal, perceiving ambiguous actions in line with the information found in the context. This builds on findings that the perceived speed of actions can influence the perception of duration (Orgs et al., 2011, 2013); predictions about rate and duration can in turn affect the identification of those actions.

Of course, just because participants adapt to rate does not necessarily imply that viewers are making predictions, per se. Consider a person standing in a dark room. After a time, a light is turned on. That light will seem intense, more intense than it would be without that sharp contrast from darkness. However, that is not to say that the brightness of the light was predicted in some way, nor that any predictions were affecting the subjective perception of the light. The same might be occurring here; viewers may perceive the short version of a final action simply by contrast with previously occurring information. Similar aftereffects are present when considering visual perception of temporal intervals (Heron et al., 2012; Li et al., 2015; Walker et al., 1981); the duration of later-occurring simple visual events is coloured by the duration of previously occurring ones. Distinguishing a contrast account from a prediction account requires a different type of paradigm; for example, one that requires viewers to label the continuation of a series of actions as congruent or incongruent with previous context (Graf et al., 2007; Schütz-Bosbach & Prinz, 2007).

Viewers may also be engaging in backward inference, with recollections of the final critical actions being perceived in line with later-generated schema (Papenmeier et al., 2019). However, on that last point, it is noteworthy that viewers were extrapolating from the preceding context without any subsequently occurring visual input. The sequence stopped after the point of ambiguity. Sequences where subsequent context was removed are precisely the contexts in which Papenmeier et al. (2019) found that no inference took place. To the extent that backward inference played a role in the responses here, these inferences would have had to occur between the final action and time at which participants responded on their keyboard, without any additional visual input. A future experiment could examine the influence of subsequent context on viewers' perceptions of the touchscreen actions. Backward inference and contrast could be bolstered or hindered by subsequent context in a way not anticipated by solely predictive accounts.

It is an open question whether these results would scale to different aspects of perception. At the very least, viewers' adaptation to the rate of events bears a striking similarity to their adaptation to rate in the auditory domain. People hearing a sequence of non-linguistic tones perceive the duration of subsequent-occurring tones based on the tempo of previous ones (McAuley & Miller, 2007). In speech perception, as with the swipe action used here, context rate can change the identity of a speech sound being produced; it can lead a "w" sound (relatively long) to be perceived as a "b" (relatively short) (Miller & Liberman, 1979). And, as with the two separate taps being perceived as a double tap, it can also lead two instances to be seen as one; "minor or child" can become just "minor child" by slowing down the context (Dilley & Pitt, 2010). Rate context effects could be a general part of the way that humans perceive the world regardless of modality. This suggests that examining the interplay between the perception of rate in speech and in action perception may be an intriguing cross-modal avenue of future investigation into the domain specificity of timing mechanisms (Carrozzo et al., 2010; Carrozzo & Lacquaniti, 2013; Pozzo et al., 2006). For instance, the duration of simultaneously occurring beeps can influence the perceived duration of flashes of light, although this was not true when the two types of stimuli were presented asynchronously (Romei et al., 2011).

Other future directions could stay firmly rooted in visual perception. The current study used a single actor performing a series of actions that were performed as a fluent sequence before the critical region. In speech, it has been shown that some rate adaptation occurs across contexts; for example, a female voice speaking slowly can affect the perception of a subsequent male speaker (Newman & Sawusch, 2009). Future studies could examine if rate adaptation could be triggered by actions with a clear discontinuity from the final sequence. A simple example would be to have the preceding context in the current study depict an actor completely different from the actor used for the ambiguous actions. Another such discontinuity would result from the context actions being unrelated to the critically ambiguous ones. For example, a sequence of dance moves preceding the ambiguous touchscreen actions would have little plausible connection to the touchscreen actions, making it an interesting test case for the influence of rate independent of semantic congruity. It also remains to be seen whether these effects are limited to animate action sequences or whether effects could also be seen in sequences of inanimate motions, including sequences of abstract shapes or inanimate, real-world objects.

To conclude, viewers can adapt to the rate of preceding actions in the context when disambiguating ambiguous actions. They can do this both to decide how to segment events as well as to determine the identity of individual events. This is in line with predictive and contrast accounts of visual perception and can occur even when immediate context is held constant. Although resembling effects is found in the auditory domain, further exploration is necessary to determine the extent to which this process can extend across modalities.

Acknowledgements

The authors would like to thank Scott Kaplowitz and the members of the Language Development Lab for their assistance in running this experiment, as well as Bob Slevc, Ellen Lau, Catherine Carr, and Gerry Altmann for their feedback.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:

This work was supported by a grant from the University of Maryland (UMD), College of Behavioral and Social Sciences to C.C.H.; a Doctoral Dissertation Improvement grant from the National Science Foundation (BCS 1650791) for W.J.I., C.C.H., and R.S.N.; and a SBE Postdoctoral Research Fellowship for C.C.H. (SMA 1714858). Portions of these experiments were used in the unpublished PhD dissertation of C.C.H. at UMD and were presented at the 58th Annual Meeting of the Psychonomics Society in Vancouver, BC. Raw data, to the extent granted by the participants in these studies, and sample stimuli will be shared online using the Open Science Framework.

ORCID iD

Christopher C Heffner https://orcid.org/0000-0002-5753-4543

Notes

1. Due to experimenter error, "twist" was used for this action during the instructions at the beginning of the study, while "rotate" was given during the experimental trials themselves.
2. This suite of experiments was designed with the idea that "press" actions would also be ambiguous with "tap" actions. Experiment 2 included many trials that ended in "press" actions in line with this expectation. However, "press" actions did not end up being ambiguous with "tap" actions and were therefore not used as critical items in Experiments 1 and 3.
3. This meant that a handful of tap actions could be seen as having six actions in the sequence if the last two elements were segmented as a double tap. This may have led to differences in the baseline likelihood of seeing two taps versus a double tap, given the other regularities in the signal, but likely would not interact with rate.

References

Arnold, D. H., & Yarrow, K. (2011). Temporal recalibration of vision. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1705), 535–538. https://doi.org/10.1098/rspb.2010.1396

Baayen, R. H., Bates, D. M., Kliegl, R., & Vasishth, S. (2015). *RePsychLing: Data sets from Psychology and Linguistics experiments*. https://github.com/dmbates/RePsychLing

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2016). *lme4: Linear mixed-effects models using Eigen and S4*. http://cran.r-project.org/package=lme4

Bertamini, M. (1993). Memory for position and dynamic representations. *Memory & Cognition*, *21*(4), 449–457.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73. https://doi.org/10.1146/annurev.psych.57.102904.190152

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), e9.

Carrozzo, M., & Lacquaniti, F. (2013). Effects of speeding up or slowing down animate or inanimate motions on timing. *Experimental Brain Research*, *224*, 581–590. https://doi.org/10.1007/s00221-012-3338-7

Carrozzo, M., Moscatelli, A., & Lacquaniti, F. (2010). Tempo Rubato: Animacy speeds up time in the brain. *PLOS ONE*, *5*(12), e15638. https://doi.org/10.1371/journal.pone.0015638

Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, *3*(5), 172–178.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Elsner, C., Falck-Ytter, T., & Gredebäck, G. (2012). Humans anticipate the goal of other people's point-light actions. *Frontiers in Psychology*, *3*, e120. https://doi.org/10.3389/fpsyg.2012.00120

Faber, M., & Gennari, S. P. (2015). In search of lost time: Reconstructing the unfolding of events from memory. *Cognition*, *143*, 193–202. https://doi.org/10.1016/j.cognition.2015.06.014

Finke, R. A., & Freyd, J. J. (1985). Transformations of visual memory induced by implied motions of pattern elements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 780–794.

Fowler, C. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, *88*(3), 1236–1249. https://doi.org/10.1121/1.399701

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 126–132.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage*, *36*(Suppl. 2), T22–T32. https://doi.org/10.1016/j.neuroimage.2007.03.017

Green, P., MacLeod, C., & Alday, P. (2016). simr: Power analysis for generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.

Hagura, N., Kanai, R., Orgs, G., & Haggard, P. (2012). Ready steady slow: Action preparation slows the subjective passage of time. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1746), 4399–4406. https://doi.org/10.1098/rspb.2012.1339

Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N. W., McGraw, P. V., & Whitaker, D. (2012). Duration channels mediate human time perception. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 690–698. https://doi.org/10.1098/rspb.2011.1131

Hudson, M., Nicholson, T., Simpson, W. A., Ellis, R., & Bach, P. (2016). One step ahead: The perceived kinematics of others' actions are biased toward expected goals. *Journal of Experimental Psychology: General*, *145*(1), 1–7.

Jacobs, A., Pinto, J., & Shiffrar, M. (2004). Experience, context, and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and*

Performance, *30*(5), 822–835. https://doi.org/10.1037/0096-1523.30.5.822

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology*, *16*, 472–479. https://doi.org/10.1016/j.cub.2006.01.032

Kaneko, S., & Murakami, I. (2009). Perceived duration of visual motion increases with speed. *Journal of Vision*, *9*(7), e14. https://doi.org/10.1167/9.7.14

Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile-visual asynchronous stimuli. *Neuroscience Letters*, *430*, 130–134. https://doi.org/10.1016/j.neulet.2007.10.044

Li, B., Yuan, X., Chen, Y., Liu, P., & Huang, X. (2015). Visual duration aftereffect is position invariant. *Frontiers in Psychology*, *6*, e1536. https://doi.org/10.3389/fpsyg.2015.01536

Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 549–554. https://doi.org/10.1037/a0027228

Magliano, J. P., Kopp, K., Higgs, K., & Rapp, D. N. (2017). Filling in the gaps: Memory implications for inferring missing content in graphic narratives. *Discourse Processes*, *54*(8), 569–582. https://doi.org/10.1080/0163853X.2015.1136870

Martel, L., Bidet-Ildei, C., & Coello, Y. (2011). Anticipating the terminal position of an observed action: Effect of kinematic, structural, and identity information. *Visual Cognition*, *19*(6), 785–798. https://doi.org/10.1080/13506285.2011.587847

McAuley, J. D., & Miller, N. S. (2007). Picking up the pace: Effects of global temporal context on sensitivity to the tempo of auditory sequences. *Perception & Psychophysics*, *69*(5), 709–718.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*(6), 457–465. https://doi.org/10.3758/BF03213823

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, *37*(1), 46–65. https://doi.org/10.1016/j.wocn.2008.09.001.Perceptual

Noel, J.-P., De Niear, M., Van der Burg, E., & Wallace, M. T. (2016). Audiovisual simultaneity judgment and rapid recalibration throughout the lifespan. *PLOS ONE*, *11*(8), e0161698. https://doi.org/10.1371/journal.pone.0161698

Orgs, G., Bestmann, S., Schuur, F., & Haggard, P. (2011). From body form to biological motion: The apparent velocity of human movement biases subjective time. *Psychological Science*, *22*(6), 712–717. https://doi.org/10.1177/0956797611406446

Orgs, G., & Haggard, P. (2011). Temporal binding during apparent movement of the human body. *Visual Cognition*, *19*(7), 833–845. https://doi.org/10.1080/13506285.2011.598481

Orgs, G., Kirsch, L., & Haggard, P. (2013). Time perception during apparent biological motion reflects subjective speed of movement, not objective rate of visual stimulation. *Experimental Brain Research*, *227*, 223–229. https://doi.org/10.1007/s00221-013-3502-8

Ortega, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2012). Flicker adaptation of low-level cortical visual neurons contributes to temporal dilation. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1380–1389. https://doi.org/10.1037/a0029495

Papenmeier, F., Brockhoff, A., & Huff, M. (2019). Filling the gap despite full attention: The role of fast backward inferences for event completion. *Cognitive Research: Principles and Implications*, *4*, e3.

Parkinson, J., Springer, A., & Prinz, W. (2011). Can you see me in the snow? Action simulation aids the detection of visually degraded human motion. *Quarterly Journal of Experimental Psychology*, *64*(8), 1463–1472. https://doi.org/10.1080/17470218.2011.594895

Pavlova, M., Staudt, M., Sokolov, A., Birbaumer, N., & Krägeloh-Mann, I. (2003). Perception and production of biological movement in patients with early periventricular brain lesions. *Brain*, *126*(3), 692–701. https://doi.org/10.1093/brain/awg062

Pozzo, T., Papaxanthis, C., Petit, J. L., Schweighofer, N., & Stucchi, N. (2006). Kinematic features of movement tunes perception and action coupling. *Behavioural Brain Research*, *169*(1), 75–82. https://doi.org/10.1016/j.bbr.2005.12.005

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *Quarterly Journal of Experimental Psychology, B*, *52*(3), 273–286.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Frontiers in Psychology*, *2*, e215. https://doi.org/10.3389/fpsyg.2011.00215

Schütz-Bosbach, S., & Prinz, W. (2007). Prospective coding in event representation. *Cognitive Processing*, *8*(2), 93–102. https://doi.org/10.1007/s10339-007-0167-x

Sgouramani, H., & Vatakis, A. (2014). "Flash" dance: How speed modulates perceived duration in dancers and non-dancers. *Acta Psychologica*, *147*, 17–24. https://doi.org/10.1016/j.actpsy.2013.06.009

Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, *4*(6), 379–384.

Simon, D. M., Noel, J.-P., & Wallace, M. T. (2017). Event Related Potentials Index rapid recalibration to audiovisual temporal asynchrony. *Frontiers in Integrative Neuroscience*, *11*, e8. https://doi.org/10.3389/fnint.2017.00008

Thornton, I. M., & Hayes, A. E. (2004). Anticipating action in complex scenes. *Visual Cognition*, *11*(2/3), 341–371. https://doi.org/10.1080/13506280344000374

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *Journal of Neuroscience*, *33*(37), 14633–14637. https://doi.org/10.1523/JNEUROSCI.1182-13.2013

Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific Reports*, *5*, 14526. https://doi.org/10.1038/srep14526

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*(1/2), 217–232. https://doi.org/10.1080/13506280143000403

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Walker, J. T., Irion, A. L., & Gordon, D. G. (1981). Simple and contingent aftereffects of perceived duration in vision and audition. *Perception & Psychophysics*, *29*(5), 475–486.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008. https://doi.org/10.1016/j.cogsci.2004.06.003

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, *1076*(1), 150–162. https://doi.org/10.1016/j.brainres.2005.12.122

Zhou, B., Yang, S., Mao, L., & Han, S. (2014). Visual feature processing in the early visual cortex affects duration perception. *Journal of Experimental Psychology: General*, *143*(5), 1893–1902.