Int. J. Human-Computer Studies 72 (2014) 358–365

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

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

Predictivity of system delays shortens human response time ☆

Roland Thomaschke [a,b,*], Carola Haering [b]

[a] Institut für Psychologie, Universität Regensburg, Universitätstraße 31, 93053 Regensburg, Germany
[b] Institut für Psychologie, Universität Würzburg, Röntgenring 11, 97070 Würzburg, Germany

ARTICLE INFO

Article history:
Received 3 September 2012
Received in revised form
23 October 2013
Accepted 6 December 2013
Communicated by P. Mulholland
Available online 16 December 2013

Keywords:
System delay
Human computer interaction
Predictivity
Predictability

ABSTRACT

System delays considerably affect users' experience and performance. Research on the psychological effects of system delays has focused on delay length and variability. We introduce delay predictivity as a new factor profoundly affecting user performance. A system delay is predictive when its duration is informative about the nature of consecutive interaction events. We report an experiment ($N=122$) where short delays were differently distributed across two alternative target stimuli in a choice response task. We manipulated variability and predictivity of delays. For one group of participants the delays were of constant duration. For three other groups the delays were variable, but differed in predictivity. They were either non-predictive, probabilistically predictive (they predicted the targets with a probability of 0.8), or deterministically predictive. Performance with constant delays was superior to performance with variable non-predictive or with probabilistically predictive delays. Surprisingly, participants with deterministically predictive delays outperformed participants in all other groups. This has important implications for interface design, whenever there is some degree of freedom in scheduling system delays. Best performance is achieved with predictive delays, but only when deterministic predictivity can be achieved. Otherwise, constant delays are to be preferred over variable ones.

1. Introduction

When interacting with a computer, users often encounter waiting times between their input and the computer's response. These delays are commonly referred to as system delays (Selvidge et al., 2002; Szameitat et al., 2009) or system response times (Dabrowski and Munson, 2011; Schleifer and Amick, 1989). System delays are, on one hand, caused by constant properties of the system such as processing speed, network bandwidth or the complexity of the requested computation. On the other hand a number of transient factors influence system delays, such as network congestion, background processes, or a variety of other factors (Seow, 2008). Research on Human Computer Interaction (HCI) has shown that system delays can enormously influence users' experience and performance (Ceaparu et al., 2004; Nah, 2004; Thum et al., 1995). Although, due to a tremendous increase in computational processing speed, system delays are negligible in some contemporary HCI interfaces, they are still a major cause for users' discomfort and low performance in others (e.g., the Internet, see also Rose et al., 2009; Seneler et al., 2009). Many recent studies have investigated how the negative effects of delays can be managed, or (if possible) avoided by interface design (Branaghan and Sanchez, 2009; Galletta et al., 2006; Krejcar, 2009).

Two important factors determining the effects of delays on users' experience and performance are the delays' lengths, and their variability (Kuhmann, 1989; Kuhmann et al., 1987; Schaefer, 1990). Before introducing a third factor – predictivity – we briefly review previous literature on delay length and variability.

1.1. The length of system delays

There is an almost universal consensus in the literature that long waiting times are detrimental to users' performance and satisfaction (Martin and Corl, 1986; Schaefer, 1990; Seow, 2008; Simoens et al., 2011). Particularly, long waiting time in internet applications do considerably affect performance and lead to user frustration. Thus, loading time is a major issue in quality of service in the context of internet applications (Liaw and Huang, 2006). Even in domains with much shorter delays, like computer games, delays have been shown to negatively affect performance (Szameitat et al., 2009). Occasionally, performance improvements by lengthening of delays have been described (e.g., Barber et al., 1983; Sellier and Chattopadhyay, 2009). These instances seem, however, to be restricted to contexts where duration of a process signals trustworthiness, like, for example, online-payment mechanisms.

☆This paper has been recommended for acceptance by P. Mulholland.

* Corresponding author at: Institut für Psychologie, Universität Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany. Tel.: +49 941 943 3776; fax: +49 941 943 1995.

E-mail addresses: Roland.Thomaschke@psychologie.uni-regensburg.de (R. Thomaschke), haering@psychologie.uni-wuerzburg.de (C. Haering).

1071-5819/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijhcs.2013.12.004

Thus, designers should, if possible, reduce the length of system delays, in order to increase user performance and satisfaction. These findings are corroborated by numerous studies in cognitive psychology, showing that response times to target stimuli increase with the length of preceding warning intervals of constant duration, except for very short (<300 ms) intervals (e.g., Leonhard et al., 2012; see Los and Schut, 2008; Müller-Gethmann et al., 2003, for reviews). These studies typically apply the *foreperiod paradigm* (Niemi and Näätänen, 1981). In this paradigm a target stimulus is preceded by a task irrelevant warning stimulus. The duration between warning and target stimulus – referred to as *foreperiod* – systematically affects performance (Rolke, 2008; Rolke et al., 2007; Rolke and Hofmann, 2007; Seibold et al., 2011; Seibold et al., 2011). However, when durations vary randomly between trials, performance *increases* with foreperiod duration (Los and Horoufchin, 2011; Steinborn and Langner, 2011; Steinborn et al., 2010; Steinborn et al., 2008; Steinborn et al., 2009). Yet, overall, responses are on average slower for variable than for constant foreperiods (Cardoso-Leite et al., 2009; Los et al., 2001; Mattes and Ulrich, 1997).

1.2. Variability of delays

The findings concerning behavioral effects from variable foreperiods in basic cognitive psychology have been confirmed in applied research with human–computer interfaces. The variability of delays the user encounters in human–machine interaction can considerably affect performance and satisfaction. System delays are referred to as *constant* when all system responses follow the preceding user input after one and the same time interval. System delays are referred to as *variable,* when the user is confronted with more than one possible delay duration. Variability can come in different degrees. Roast (1998) has defined the degree of variability as the span between the shortest and the longest possible interval duration (see also Fischer et al., 2005). It is a well-established finding in basic human performance research that choice responses are on average faster after constant than after variable delays (Cardoso-Leite et al., 2009; Wundt, 1874, see above). Since the early days of ergonomic research, this finding has been validated in several studies in human–machine interaction (Awramoff, 1903; Weber et al., 2013). HCI research has shown that increased variability has detrimental effects on user satisfaction (Fischer et al., 2005) and performance (Weber et al., 2013). Weber et al. (2013), for example, manipulated delay variability in an E-Mail program. Users' response latencies significantly increased with increasing delay variability.

1.2.1. Reducing variability: shortening and lengthening

As delay variability has negative effects on user experience and performance, designers should attempt to minimize variability in HCI interfaces. There are several ways to reduce variability. One option is obviously the reduction of extraordinarily long delays, like, for example, internet-download times. In order to reduce variability, such reduction must be specific to long delays, in the sense that already short delays are not also shortened (Roast, 1998). As the delays are caused by factors inherent in the system, delay reduction requires some kind of technical optimization of the computational process that causes long delays. Such technical optimizations, however, are beyond the scope of HCI interface design, and are, thus, not the focus of the present study.

A technically much less demanding way of reducing variability is the selective lengthening of short delays. This is the approach taken by Weber et al. (2013). In that study, for one group of participants, interaction with an email program was unpredictably interrupted by delays of 7 different durations. In another group, 5 of the possible durations were lengthened in a way that all delays could now have only 2 different lengths. The selective lengthening of delays considerably improved participants' performance. Likewise, Sellier and Chattopadhyay (2009) suggested to selectively add delays to unusually short web-page loading times to avoid the impression that "something is not right" with the web pages. Selective lengthening has the advantage that no sophisticated technical improvement is required. No computational process must be optimized. Delays must only be added in appropriate places.

It has, however, the major disadvantage that it also prolongs the total interaction time with the system. This issue becomes particularly problematic with regard to user performance and user frustration. An important reason to reduce variability is to improve user performance in the sense of speeding up users' responses (see Szameitat et al., 2009; Weber et al., 2013). It is, however, unlikely that the delays added to reduce variability will be compensated by the times saved by shorter user response latencies (though there are presently no systematic investigations on this issue). However, in Weber et al.'s study, due to the longer delays, total interaction time was longer in the low variability condition, although users' response latencies were reduced. Nevertheless, user satisfaction was not decreased in the condition with on average longer waiting times.

1.2.2. Reducing variability: scheduling

Another means for reducing variability is scheduling of delays. As described above, changing the obtainable system speed per se is beyond the scope of interaction design. However, scheduling enables interaction designers to speed up system response times by optimizing the use of processing power (Blazewicz et al., 2007). Scheduling requires that there are at least some degrees of freedom concerning the point in time when an interaction has to take place during the computational process. Scheduling is obviously not possible when the processing capacities of the system are at any time exclusively devoted to processing one input of one individual user. This is, however, the model implicitly or explicitly assumed by most traditional models informing temporal variability research in HCI (e.g., Roast, 1998). Consequently, scheduling has not been considered as an option to reduce variability.

Most modern computer systems are, however, not covered by those models. Due to the growing application of parallel computing, it is often the case that different processes or different users share a single processor or a set of processors. In such scenarios the need for some kind of scheduling emerges (Szameitat et al., 2009). The interface designer has some degree of choice how to distribute processing time over interaction events.

For example, in many programs' download and installation procedures, dialogs with the user are scheduled parallel to the download. Users provide information about the installation path, program settings etc. while the program is already downloading. This renders the system's delays less variable compared to situations with one long delay during the download and several almost instantaneous dialog interactions before or after (see Seow, 2008). Another example is an algorithm, developed by Pons (2006), which reschedules processing capacity from fast loading to long loading web pages, in order to reduce delay variability.

Scheduling combines the advantages of shortening and lengthening delays, which were discussed above. On the one hand, it makes system delays less variable without making computational processing technically faster. On the other hand, rescheduling avoids adding empty delays during which no processing takes place. Thus, variability can be reduced without artificially lengthening the total interaction time. Scheduling allows a system designer to homogenize intervals (e.g., by separating long delays and uniting short ones), and to also manipulate regularity between delays and interaction-events.

1.3. Predictivity of system delays

Previous studies that investigated the effects of delays on user performance did not take into account the relations between system delays and the types of system response. Yet in real life human computer interactions, system delays and system responses stand in various systematic prediction-relations. In this paper we introduce a systematic conceptual classification of such relations, and at present the first empirical test of these relations' effectiveness on user performance.

In many interaction contexts the duration of the delay preceding a system's response is informative about the nature of the system's response. When, for example in a search interface, error messages are usually displayed faster than hits, the system delays predict the type of the following system response. We refer to delays that are informative about the next system response as *predictive* delays.

Predictivity of a delay can be contrasted with *predictability*. We refer to a delay as predictable, when its duration can be predicted by previous events – be it a preceding system response, or a user response. Delays are predictable when, for example, the user's responses are correlated with consecutive delays, like when complex requests are followed by longer processing time than simple requests. Predictivity and predictability are inverse concepts: predictive delays predict, and predictable delays are predicted. Basic research on temporal expectation has shown that humans are sensitive to interval predictivity (Thomaschke and Dreisbach, 2013; Wagener and Hoffmann, 2010; Wendt and Kiesel, 2011) as well as to interval predictability (Correa et al., 2004; Coull and Nobre, 1998; Haering and Kiesel, 2012; Kingstone, 1992; Lawrence and Klein, 2013).

In the present paper we focus on the behavioral effects of predictivity. While variability is a direct property of the set of delays itself, independent of how the delays are distributed across system's responses, predictivity is a property of the *relation* of the set of delays to the system's responses. However, predictivity requires at least some degree of variability. Thus, constant delays cannot be predictive, by definition. Except for that restriction, variability and predictivity of delays can be manipulated independently of each other, because they refer to different aspects of a computational system.

Predictivity can also be manipulated in different degrees. We define the degree of predictivity of a given set of delays, as the strength of correlation between the delays and the types of system responses immediately following the delays. Consider, for example, a system with two possible system responses (e.g., error, or success), and two possible delays.[1] When, these delays are correlated by 0.5 with events, both responses are equally likely after both delays. The delays are *non-predictive* in this case. When the correlation is higher, for example 0.8, then the length of the delays carries some probabilistic information about the following event. After the shorter delay, one of the responses is more likely, after the longer delay, the other one is more likely. In such a situation, we refer to the set of delays as *probabilistically* predictive. When one of the delays is always followed by one and the same system's response, and the other one is always followed by the other response, then the correlation is 1, and the set of delays is referred to as *deterministically* predictive.

Many real computer systems include delays which are to some degree predictive. For example, the probability that a Web page will load successfully decreases continuously after navigating to the URL, until an error message becomes more likely (Seow, 2008). A study by Shahar et al. (2012) has revealed that users exploit system delays to infer the functional state of a computer system.

Previously it has been demonstrated that humans are sensitive to predictivity (Roberts et al., 2011; Thomaschke et al., 2011a, 2011b; Watanabe et al., 2008, see above). However, little is known about the effects of delay predictivity on user performance. The present study investigates whether probabilistically or deterministically predictive delays lead to faster response times. As total delay time is kept constant over different predictivity conditions (it is just scheduled differently among system's responses), faster user response times would mean shorter over alltask completion times.

1.4. Overview of the present study

In the present study we investigate for the first time whether predictivity of system delays speeds up users' responses. Furthermore we investigate how potential gains in users' response speed by delay predictivity are related to the known detrimental effects of delay variability.

As this is the first step in researching delay-predictivity effects on user performance, we have chosen one of the simplest interaction tasks in HCI research: in a binary choice reaction task, participants had to classify easily identifiable stimuli by key press responses. We manipulated the variability and predictivity of delays between groups of participants.

Two groups with constant delays were compared to three groups with variable delays. According to previous research on delay variability, constant delays should lead to faster responses than variable delays (Cardoso-Leite et al., 2009; Roast, 1998). The three variable groups differed in predictivity. In a non-predictive group, two possible delays were equally distributed over both target stimuli. In a probabilistically predictive group, the same delays predicted the upcoming target with 80% probability. In a deterministically predictive group, delays were deterministically coupled with targets. If predictivity shortens human response time, we would observe better performance in the two predictive groups than in the non-predictive group. When the degree of predictivity matters, we would observe better performance in the deterministically, compared to the probabilistically, predictive group.

2. Method

2.1. Design

We devised a binary forced choice paradigm with warning intervals – here referred to as delays – between target stimuli. Two possible delays (400 ms and 1000 ms) were differently distributed over target stimuli (see Table 1) for different groups of participants. One group experienced only the 400 ms delay throughout the experiment, while another group experienced only the 1000 ms delay. In the following we refer to both groups together as the *constant* group. The same delays – 400 ms and 1000 ms – have been used for the three variable delay groups. Each of the delays appeared in half of the trials, in these groups. The sequence of delays was randomly determined. The variable groups differed, however, with regard to the frequency of delay-target *combinations*. In the non-predictive group, each target was equally often paired with the long and with the short delay. In the probabilistically predictive group, one target appeared frequently (at 80% of its occurrences) after the short delay, while the other target was frequent after the long delay. In the deterministically predictive group, one of the targets appeared only after the short delay, while the other target appeared only after the long delay. The

[1] For the sake of simplicity we deal only with binary sets of delays in this study. But the concepts introduced here can, in principle, be extended to each number of possible delays. Whether the empirical findings of the present study would also generalize to non-binary task scenarios is, however, an open question (see discussion section).

Table 1
Number of trials in each group and experimental condition.

Experimental group	Delay (ms)	Target 1	Target 2	Sum (%)
Constant group (short) (N=25)	400	600	600	100
	1000	0	0	0
Constant group (long) (N=25)	400	0	0	0
	1000	600	600	100
Variable non-predictive group (N=24)	400	300	300	50
	1000	300	300	50
Variable probabilistically predictive group (N=24)	400	480	120	50
	1000	120	480	50
Variable deterministically predictive group (N=24)	400	600	0	50
	1000	0	600	50
Sum (%)		50%	50%	

association between target and delay was counterbalanced between participants.

2.2. Participants

Participants were students of the University of Würzburg. They were paid 6 €, or received course credit. Participants were naïve as to the purpose of the experiment and were randomly allocated to experimental groups. 102 from 122 participants were female. The mean age was 22.43, SD=3.2. They had normal or corrected to normal vision. See Table 1 for allocation of participants to groups.

2.3. Apparatus and stimuli

The experiment was run on a standard PC, equipped with a 17″ CRT monitor. The delays were triggered by pressing the left or right mouse button with the right hand. Responses to the target were given on two response keys with the left hand. Target stimuli were a black star or a black triangle (both 1.5 cm × 1.5 cm) on a white background. The fixation cross was a black "+" sign in the font Arial (0.5 cm × 0.5 cm).

2.4. Procedure

Each trial started with presentation of the fixation cross. Participants were to press any mouse button whenever they felt ready to. Immediately after pressing the button the font-weight of the fixation cross was changed from regular to bold. After a delay of either 400 ms or 1000 ms, the fixation cross was substituted by the target stimulus. Participants had to respond as fast as possible by pressing the left or right response key according to the target stimulus. The assignment of target stimuli (star or triangle) to the left or right key was counterbalanced across participants. With the response the target disappeared. When participants responded timely and correctly, the next trial started after 800 ms with appearance of the fixation cross. When participants chose the wrong key an error message ("Falsche Taste!", German for "Wrong key!") appeared accompanied by an error sound. When participants did not respond to the target within 700 ms, the error message "Bitte schneller!" (German for "Faster, please!") appeared, accompanied by an error sound.

Four groups of participants differed in variability and predictivity of delays (see above). The experiment consisted of 12 main blocks of 100 trials each. In the main blocks the order of targets and delays randomly varied from trial to trial except in the constant group, who only experienced one delay and varying targets. The experimental blocks were preceded by 2 short baseline blocks, which served to calculate decrement scores (see below). The baseline blocks comprised of 24 trials each. The delays in both baseline blocks were constant. In one block, only the short delay was used, while only the long delay was presented in the other baseline block. The order of baseline blocks was counterbalanced between participants. There were self paced breaks between baseline and between experimental blocks. The total duration of the experiment was about 50 min for the constant group with short delays, about 55 min for the variable groups, and about 60 min for the constant group with long delays.

3. Results

3.1. Data preparation

3.1.1. Calculation of the individual RT baseline

The participants with long constant delays and with short constant delays were collapsed into one experimental group, because the distribution of delays over events was the same for both groups.

As outlined in the methods section, we calculated a RT baseline for each participant from the two short baseline blocks. The first four trials of each baseline block were excluded. For the remaining 20 trials, the percentages of errors (wrong or too late responses) were calculated. A one-way ANOVA confirmed that there were no group differences regarding error rates in the baseline blocks, $F(3,118)=1.724$, $p=.167$. Error-trials and trials following error-trials were removed. Per participant and per baseline block, there were on average 17.69 trials (SD=2.47; range: 8–20) left from the originally 24 trials. From these remaining trials, the average response speed over both baseline blocks has been calculated as a RT baseline for each participant. The average RT baselines for the constant delay group, $M=396.86$ ms, SD=53.88, the non-predictive group, $M=397.64$ ms, SD=49.54, the probabilistically predictive group, $M=391.49$ ms, SD=40.69, and the deterministic group, $M=402.62$ ms, SD=39.01, did not differ significantly from each other, $F(3,118)=0.233$, $p=.873$.

3.1.2. Preprocessing of RT data

Parts of the first experimental block might have been under the influence of the baseline phase (although it was very short). In the conditions with variable delays, this new temporal variation might have been surprising for the participants, in contrast to the group with constant delays, because all groups experienced constant delays in the short practice phase. Because this potential surprise could have confounded our comparison between groups, we excluded the first block from the analysis. We also removed the first five trials of each experimental block, to minimize potential variation from the self paced breaks between blocks.

Participants committed on average 4.588% errors, SD=4.31. As the error scores did not significantly differ between conditions, $F(3,118)=0.351$, $p=.788$, we removed error trials from the RT analysis. We also removed trials following error trials, because RTs after errors are usually strongly affected by the preceding error (Laming, 1968; Steinhauser and Kiesel, 2011).

For the RT analysis, we further removed all trials with RTs deviating from the participant's mean RT more than 2.5 SDs (as recommended by Bush et al., 1993; Whelan, 2008). These were 1.8% of all trials.

To calculate each participant's mean RT-decrement from baseline (the focus of our analysis), we simply subtracted each participant's mean RT in experimental blocks from that participant's baseline RT.

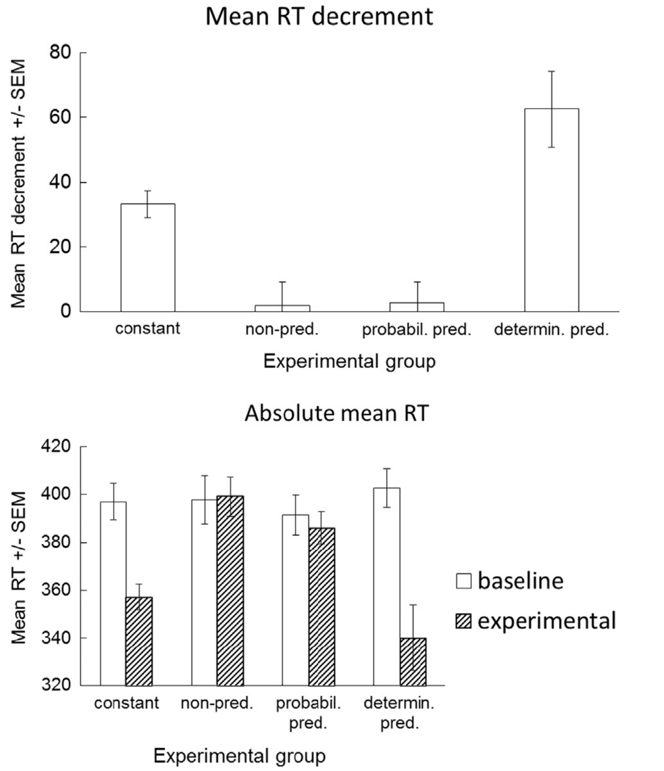

Fig. 1. Mean RT decrement from baseline (upper panel), and absolute mean RT in baseline and experimental blocks (lower panel), in dependence on experimental group. Error bars represent 1 SEM.

3.2. Effects of group on user response time

Absolute mean RTs and mean RT decrement (relative to the pre-experimental baseline), are displayed in Fig. 1. We analyzed RT decrement, because this score is not dependent on inter-subjective baseline difference in response speed. However, we always report the analogous results for absolute RT in parentheses. A one-way between subjects ANOVA showed that experimental group affected RT decrement, $F(3,118)=13.981$, $p<.001$ ($F=9.484$, $p<.001$, for absolute RT). Bonferroni corrected t-tests showed that participants in the deterministically predictive group outperformed the participants in each of the remaining three groups; against the constant delay group, M_D (mean difference)$=29.37$ ms, $p=.018$ ($M_D=17.57$ ms, $p=.708$), against the non-predictive group, $M_D=60.68$ ms, $p<.001$ ($M_D=59.56$ ms, $p<.001$), and against the probabilistically predictive group $M_D=59.91$ ms, $p<.001$ ($M_D=46.42$ ms, $p=.002$). RT reduction in the group with constant delays was significantly higher than in the non-predictive group, $M_D=31.31$ ms, $p<.009$ ($M_D=41.89$ ms, $p=.002$), and in the probabilistically predictive group, $M_D=30.54$ ms, $p=.009$ ($M_D=28.85$ ms, $p=.054$). The only non-significant difference was between the non-predictive and the probabilistically predictive group, $M_D=0.77$ ms, $p>.999$ ($M_D=13.14$ ms, $p>.999$).

An analogous ANOVA for mean error rates was not significant, $F(3,118)=0.351$, $p=.788$. None of the Bonferroni-corrected post hoc tests between conditions yielded a significance lower than $p=.999$.

For the probabilistically predictive group, we conducted an additional within-subjects analysis: by comparing absolute RTs on frequent combinations of delay and target (80% of trials) with absolute RTs on infrequent combinations (20% of trials), we tested whether participants formed temporal expectancies about targets. Mean RT was shorter for frequent combinations, $M=380$ ms, than for infrequent combinations, $M=404$ ms, $t(23)=10{,}037$, $p<0.001$.

3.3. Effects from group and delay duration on user response time

Analyses in the previous subsection allow conclusions about the overall performance gains of some delay-event distributions over others. However, in order to investigate how these gains have cognitively been produced, we also analyzed the data separately for the 400 ms and the 1000 ms delays. Note, that for the following delay-specific analyses, the constant group included only half of the participants than in the main analyses, because half of the constant group was exposed to the short delay only, and the other half to the long delay only (see Table 1). However, in the main analysis, the constant group included twice as many participants as each one of the variable groups (due to collating participants with constant 400 ms delays and with constant 1000 ms delays). Consequently, all groups in the following analysis have about the same size.

For the 400 ms delay, the ANOVA for RT decrement with the factor Experimental group was significant, $F(3,93)=8.041$, $p<.001$ ($F=5.864$, $p=.001$, for absolute RT). However, the Bonferroni-corrected post hoc tests showed a different pattern, especially for the relation between the deterministic and the other groups. The deterministically predictive group was only better than the probabilistically predictive group, $M_D=31.24$ ms, $p=.024$ ($M_D=17.76$ ms, $p=.820$). There was only a marginally significant tendency towards an advantage over the non-predictive group, $M_D=29.18$ ms, $p=.054$ ($M_D=28.06$ ms, $p=.133$), and a non-significant disadvantage compared with the constant group, $M_D=41.72$ ms, $p>.999$ ($M_D=18.74$ ms, $p=.734$). However, performance improvement in the constant delay group was again significantly higher than in the non-predictive, $M_D=59.91$ ms, $p<.001$ ($M_D=46.42$ ms, $p=.002$), and in the probabilistically predictive group, $M_D=43.78$ ms, $p=.001$ ($M_D=36.49$ ms, $p=.016$). As in the main analysis, the difference between non-predictive and the probabilistically predictive group was not significant, $M_D=2.06$ ms, $p>.999$ ($M_D=10.30$ ms, $p>.999$). An analogous ANOVA for error rates was not significant, $F(3,93)=0.538$, $p=.657$. None of the Bonferroni-corrected post hoc tests reached a corrected p value below .999.

For the probabilistically predictive group, we compared frequent with infrequent combinations at the 400 ms delay. As in the main analysis, mean RT was shorter for frequent combinations, $M=394$ ms, than for infrequent combinations, $M=405$ ms, $t(23)=3.063$, $p=0.005$.

The ANOVA for the 1000 ms delay was also significant, $F(3,93)=19.072$, $p<.001$ ($F=12.532$, $p<.001$, for absolute RT). With regard to the deterministic group, the results of the Bonferroni-corrected post hoc tests were in line with the main analysis: participants in the deterministically predictive group outperformed on average participants in all other groups; against the constant delay group, $M_D=70.75$ ms, $p<.001$ ($M_D=53.36$ ms, $p=.006$), against the non-predictive group, $M_D=91.74$ ms, $p<.001$ ($M_D=90.63$ ms, $p<.001$), and against the probabilistically predictive group, $M_D=87.68$ ms, $p<.001$ ($M_D=74.39$ ms, $p<.001$). However, deviating from the main analysis the constant group was not faster than the non-predictive, $M_D=20.99$ ms, $p=.806$ ($M_D=37.26$ ms, $p=.124$), or the probabilistically predictive group, $M_D=17.11$ ms, $p>.999$ ($M_D=21.02$ ms, $p>.999$). The probabilistically predictive and the non-predictive group did not differ significantly in their response time reduction, $M_D=3.89$ ms, $p>.999$ ($M_D=16.24$ ms, $p>.999$).

An analogous ANOVA for error rates was not significant, F $(3,93)=0.207$, $p=.892$. None of the post hoc tests reached a Bonferroni-corrected p value below .999.

For the probabilistically predictive group, we compared frequent with infrequent combinations also at the 1000 ms delay. As with 400 ms, mean RT was shorter for frequent combinations,

$M=365$ ms, than for infrequent combinations, $M=402$ ms, $t(23)=6.849$, $p<0.001$.

4. Discussion

4.1. Summary of results

We have compared four groups of participants, performing a forced choice task interrupted by system delays. We manipulated the variability and predictivity of the delays. One constant-delay group (with no variability and, hence, no predictivity) was compared with three variable delay groups, differing only in the degree of delay predictivity (non-predictive, probabilistically predictive and deterministically predictive). Our study had three purposes. The first aim was to confirm previous findings that delay variability has a detrimental effect on user performance (measured in RT). Our second aim was to investigate, for variable delay conditions, whether predictivity of delays has a potentially positive effect on user performance. Third, we wanted to elucidate how such potential positive effects would cognitively emerge by determining at which delay duration the effects are produced.

We confirmed the performance advantage for constant delays over variable delays, evident from previous literature (Cardoso-Leite et al., 2009; Wundt, 1874). Performance in the constant delay group was clearly superior to performance in the non-predictive variable delay group.

With regard to predictivity, probabilistic predictivity did not yield significant performance gain. Although participants formed temporal expectancies in the probabilistically predictive group, evidenced by faster responses to temporally expected than to temporally unexpected targets (see also Thomaschke et al., 2011; Wagener and Hoffmann, 2010), this did not result in an overall improved performance. The performance gain by temporally expected trials was compensated by performance losses in temporally unexpected trials. However, with deterministic predictivity performance was significantly better than in all other groups, including the constant delay group. This means that the positive effects of deterministic predictivity outweighed the negative effects of variability.

Separate analyses for both delay durations revealed that the performance advantage of the deterministic group over all other groups were primarily due to the long 1000 ms delay. At the short 400 ms delay, responses with deterministically predictive delays were only faster than responses with probabilistically predictive delays, but did not significantly differ from the other groups.

4.2. Practical implication

One apparent application of our results lies in the evaluation of existing computer human interfaces. When interfaces differ mainly in the distribution of delay time, the interface with constant delays is to be preferred over interfaces with variable delays. This holds, however, only when the variable delays are not deterministically predictive: an interface with variable delays where the different interaction events are always preceded by individual characteristic delays is to be preferred over an interface with constant delays. This arrangement leads to fastest user response times, with equal overall system delay times.

Our results can, however, also serve as a guideline in designing interfaces, in order to decrease overall task completion time by reduction of user response time via scheduling of system delays. Scheduling refers to systematically allocating system processing time to different tasks. Most current computer systems apply some form of scheduling (Blazewicz et al., 2007). Scheduling has to integrate several constraints. Some are related to the system (e.g., limited availability of parallel processors at a time), others are related to needs of the user (e.g., readiness of an agent to receive the next call in a call center environment).

One type of user-related constraints regards ergonomically optimal system-delay structures. Based on previous research, the most important one of these constraints for scheduling algorithms was: schedule the processing resources in a way that the system-delays for each user are approximately constant (Seow, 2008). We suggest that an additional scheduling constraint should be: if you can schedule the processing resources in a way that system delays for each user are deterministically predictive, make them deterministically predictive. If only probabilistic predictivity can be achieved, do not schedule delays predictively, but rather aim for constant delays.

When by means of scheduling variable system delays can be allocated deterministically to distinct system responses, without the involvement of additional delays, the average reduction in user response time per interaction event can be estimated around 60 ms (decrement score-difference between variable non-predictive and variable deterministically predictive group). For simple repetitive tasks with an average delay around 700 ms (as in the present study) and a user response time below 500 ms, this would yield an overall task time reduction of about 3 min per hour, without any overall reduction of system delays. Note, however, that the implications of our results might be restricted to binary delay-event associations, that means, to systems where two different system response classes can be deterministically scheduled to two delay durations (see next subsection).

4.3. Cognitive mechanisms

The main finding of our study might seem surprising at first sight. Deterministic predictivity leads to such a drastic reduction in response times that it even outweighs the well-established response time increase by variability. Being sure about what will happen at a certain time helps performance more than being unsure when something will happen impairs performance. However, a closer look at the distribution of performance at the individual delay durations shows how the substantial performance gain by deterministic predictivity was achieved, but it also points to a potential limitation of generalizability.

At the short 400 ms delay the benefits of deterministic predictivity are rather moderate. The condition with deterministic predictivity is on a par with constant and with the non-predictive group. It had only a significant 30 ms advantage over the probabilistically predictive condition. This suggests that participants utilize the deterministic knowledge about the short delay to some degree. Although they cannot be sure whether a signal will appear at 400 ms, they exactly know which symbol it would be at 400 ms. Thus they prepare the response to that symbol to some degree, though they might need to change their preparation when the short delay elapsed without presentation of a symbol. This tendency is, however, not significant for each between group comparison (note, that mean response times for 400 ms are numerically higher in the deterministic predictive group than in the constant group), and in magnitude much smaller than the large overall benefit of deterministic predictivity.

This benefit is almost exclusively caused by performance differences at the longer, 1000 ms, delay. At this delay participants in the deterministically predictive group can be sure when the symbol will appear and which symbol it will be. Thus, the task changes from a choice response task to a simple response task. Responses on single tasks are known to be much shorter than on choice tasks (Wundt, 1874). This is an advantage which the deterministically predictive group has over all other groups, including the constant group. Inspection of Fig. 2 shows that the

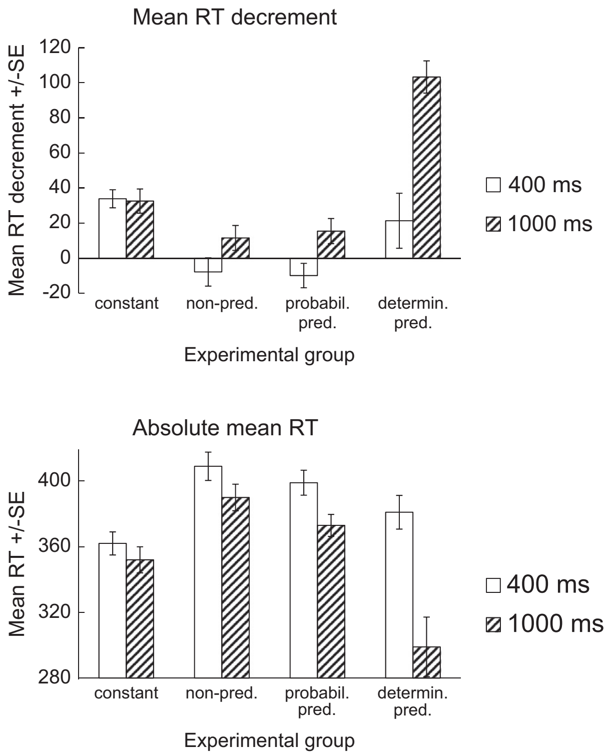

Fig. 2. Mean RT and mean RT decrement from baseline in dependence on experimental group and delay duration. Error bars represent 1 SEM.

overall benefits of predictivity can be to a large degree ascribed to this combined time and event determination at the 1000 ms delay.

This explanation does, however, point to a potential limitation of the present findings. As the benefit from predictivity seems to rely on the response certainty at the long delay, it might be that the benefit is confined to a binary response situation, or more precisely, to a situation where the longer delay is associated to only one response. In a situation where multiple (e.g., two of four) responses are scheduled deterministically to the late delay, the response set would be reduced when shorter delay had elapsed, but it would still be a choice task. Thus we speculate that the benefits would be rather similar to the ones observed at 400 ms in the present study.

4.4. Questions for further research

The present study is, to our knowledge, the first investigation into the effects of delay predictivity on computer users. Despite showing a substantial effect of delay predictivity, our investigation was focused on just one aspect of users' experience and behavior (productivity measured in user RT), and on only one type of computer user interaction (classification of system output) among many.

Concerning other possible dependent measures, several previous studies found strong increasing effects from delay duration on users' annoyance (Fischer et al., 2005; Planas and Treurniet, 1988; Williges and Williges, 1982) and detrimental effects on perceived quality of the computer system (Bhatti et al., 2000; Hoxmeier and DiCesare, 2000; Ramsey et al., 1998). One might speculate that predictivity of system delays might lead to higher perceived quality and lower annoyance with computer systems, because users can better synchronize their behavior with the system. This can, however, not be directly inferred from the present study, but would instead require a more specialized investigation.

The analysis in the present study has focused on performance instead of affective measures. Note, however, that we have not found any effects of predictivity on correctness of responses. This is in line with numerous previous studies on the effects of delay duration on performance in simple tasks (MacKenzie and Ware, 1993; Martin and Corl, 1986; O'Donnell and Draper, 1996), finding also no effects on error rates. When, on the other hand, tasks were more complex, delay duration had substantial negative effects on correctness (Barber et al., 1983; Kohlisch and Kuhmann, 1997; Schaefer, 1990; Thum et al., 1995). It would be an interesting topic for further research, whether predictivity of delay can also affect correctness when tasks are more complex than merely choice responses.

Another important issue to be investigated in relation to the present findings is determining their boundary conditions. We have provided results from an 80% and from a 100% predictive condition. Performance in the 80% condition has not shown any advantage over the 50% (i.e., non-predictive) condition, while 100% predictivity lead to a substantial improvement. It needs to be investigated whether manipulations between 80% and 100% would lead to a gradual increase of performance or whether only full deterministic predictivity would facilitate responses. Put another way, whether a small amount of deviating delay-event combinations would prevent the cognitive benefits of deterministic predictivity or not.

5. Conclusions

We have shown that variable system delays lead to longer user response times than constant delays, except when the duration of these delays predicts the type of the following system response in a deterministic fashion. In the latter case, user responses are even significantly faster than with constant delays. These results have implications for human computer interfaces where system delays can be scheduled with respect to consecutive interaction events. Delays should be variable and should be deterministically associated with individual interaction events. If this is not possible, constant delay should be preferred. This guideline would reduce total task completion time by reduction of user response time without any changes in total system delay time. It remains to be explored whether this guideline also generalizes to system delays as long as about one second, and to tasks more complex than simple binary classification.

References

Awramoff, D., 1903. Arbeit und Rhythmus: Der Einfluss des Rhythmus auf die Quantität und Qualität geistiger und körperlicher Arbeit, mit besonderer Berücksichtigung des rhythmischen Schreibens. Philos. Stud. 18, 515–562.

Barber, R.E., Henry, J., Lucas, C., 1983. System response time, operatory productivity, and job satisfaction. Commun. ACM 26, 972–986.

Bhatti, N., Bouch, A., Kuchinsky, A., 2000. Integrating user-perceived quality into web server design. Paper presented at the the 9th Word-Wide Web Conference.

Blazewicz, J., Ecker, K.H., Pesch, E., Schmidt, G., Weglarz, J., 2007. Handbook of Scheduling: From Theory to Applications. Springer, Heidelberg, New York.

Branaghan, R.J., Sanchez, C.A., 2009. Feedback preferences and impressions of waiting. Hum. Factors 51, 528–538, http://dx.doi.org/10.1177/0018720809345684.

Bush, L.K., Hess, U., Wolford, G., 1993. Transformations for within-subject designs: a Monte Carlo investigation. Psychol. Bull. 113, 566–579.

Cardoso-Leite, P., Mamassian, P., Gorea, A., 2009. Comparison of perceptual and motor latencies via anticipatory and reactive response times. Atten. Percept. Psychophys. 71, 82–94, http://dx.doi.org/10.3758/app.71.1.82.

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., Shneiderman, B., 2004. Determining causes and severity of end-user frustration. Int. J. Hum. Comput. Interact. 17, 333–356.

Correa, A., Lupiáñez, J., Milliken, B., Tudela, P., 2004. Endogenous temporal orienting of attention in detection and discrimination tasks. Percept. Psychophys. 66, 264–278, http://dx.doi.org/10.3758/BF03194878.

Coull, J.T., Nobre, A.C., 1998. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J. Neurosci. 18, 7426–7435.

Dabrowski, J., Munson, E.V., 2011. 40 years of searching for the best computer system response time. Interact. Comput. 23, 555–564.

Fischer, A.R.H., Blommaert, F.J.J., Midden, C.J.H., 2005. Monitoring and evaluation of time delay. Int. J. Hum. Comput. Interact. 19, 163–180.

Galletta, D.E., Henry, R.M., McCoy, S., Polak, P., 2006. When the wait isn't so bad: the interacting effects of website delay, familiarity, and breadth. Inf. Syst. Res. 17 (1), 20–37, http://dx.doi.org/10.1287/isre.1050.0073.

Haering, C., Kiesel, A., 2012. Time in action contexts: learning when an action effect occurs. Psychol. Res. 76, 336–344, http://dx.doi.org/10.1007/s00426-011-0341-8.

Hoxmeier, J.A., DiCesare, C., 2000. System response time and user satisfaction: an experimental study of browser-base applications. Paper presented at the Americaa Conference on Information Systems.

Kingstone, A., 1992. Combining expectancies. Q. J. Exp. Psychol. 44A, 69–104, http://dx.doi.org/10.1080/14640749208401284.

Kohlisch, O., Kuhmann, W., 1997. System response time and readiness for task execution – the optimal duration of the inter-task delays. Ergonomics 40, 265–280.

Krejcar, O., 2009. Problem solving of low data throughput on mobile devices by artefacts prebuffering. Eurasip J. Wireless Commun. Networ. 2009, 802523, http://dx.doi.org/10.1155/2009/802523.

Kuhmann, W., 1989. Experimental investigation of stress-inducing properties of system response times. Ergonomics 32, 271–280.

Kuhmann, W., Boucsein, W., Schaefer, F., Alexander, J., 1987. Experimental investigation of psychophysiological stress-reactions induced by different system response times in human–computer interaction. Ergonomics 30, 933–943.

Laming, D.R., 1968. Information Theory of Choice-Reaction Times. Academic Press, London.

Lawrence, M.A., Klein, R.M., 2013. Isolating exogenous and endogenous modes of temporal attention. J. Exp. Psychol.: Gen. 142, 560–572.

Leonhard, T., Bratzke, D., Schröter, H., Ulrich, R., 2012. Time-course analysis of temporal preparation on central processes. Psychol. Res. 76, 236–251.

Liaw, S.S., Huang, H.M., 2006. Information retrieval from the World Wide Web: a user-focused approach based on individual experience with search engines. Comput. Hum. Behav. 22, 501–517, http://dx.doi.org/10.1016/j.chb.2004.10.007.

Los, S.A., Horoufchin, H., 2011. Dissociative patterns of foreperiod effects in temporal discrimination and reaction time tasks. Q. J. Exp. Psychol. 64, 1009–1020, http://dx.doi.org/10.1080/17470218.2010.532225.

Los, S.A., Knol, D.L., Boers, R.M., 2001. The foreperiod effect revisited: conditioning as a basis for nonspecific preparation. Acta Psychol. 106, 121–145.

Los, S.A., Schut, Marcus L.J., 2008. The effective time course of preparation. Cognit. Psychol. 57, 20–55.

MacKenzie, I.S., Ware, C., 1993. Lag as a determinant of human performance in interactive systems. Paper presented at the ACM Conference on Human Factors in Computing Systems, New York.

Martin, G.L., Corl, K.G., 1986. System response time effects on user productivity. Behav. Inf. Technol. 5, 3–13.

Mattes, S., Ulrich, R., 1997. Response force is sensitive to the temporal uncerainty of response stimuli. Percept. Psychophys. 59, 1089–1097.

Müller-Gethmann, H., Ulrich, R., Rinkenauer, G., 2003. Locus of the effect of temporal preparation: evidence from the lateralized readiness potential. Psychophysiology 40, 597–611, http://dx.doi.org/10.1111/1469-8986.00061.

Nah, F.F.-H., 2004. A study on tolerable waiting time: how long are Web users willing to wait? Behav. Inf. Technol. 23, 153–163, http://dx.doi.org/10.1080/01449290410001669914.

Niemi, P., Näätänen, R., 1981. Foreperiod and simple reaction-time. Psychol. Bull. 89, 133–162, http://dx.doi.org/10.1037/0033-2909.89.1.133.

O'Donnell, P., Draper, S.W., 1996. How machine delays change user strategies. ACM SIGCHI Bull. 28 (2), 39–42.

Planas, M.A., Treurniet, W.C., 1988. The effects of feedback during delays in simulated teletext reception. Behav. Inf. Technol. 7, 183–191.

Pons, A.P., 2006. Semantic prefetching objects of slower web site pages. J. Syst. Softw. 79, 1715–1724, http://dx.doi.org/10.1016/j.jss.2006.02.005.

Ramsey, J., Barbesi, A., Preece, J., 1998. A psychological investigation of long retrieval times on the word wide web. Interact. Comput. 10, 77–86.

Roast, C., 1998. Designing for delay in interactive information retrieval. Interact. Comput. 10, 87–104.

Roberts, F., Margutti, P., Takano, S., 2011. Judgments concerning the valence of inter-turn silence across speakers of american english, italian, and japanese. Discourse Process. 48, 331–354, http://dx.doi.org/10.1080/0163853x.2011.558002.

Rolke, B., 2008. Temporal preparation facilitates perceptual identification of letters. Percept. Psychophys. 70, 1305–1313, http://dx.doi.org/10.3758/pp.70.7.1305.

Rolke, B., Bausenhart, K.M., Ulrich, R., 2007. Impaired temporal discrimination within the attentional blink. Percept. Psychophys. 69, 1295–1304.

Rolke, B., Hofmann, P., 2007. Temporal uncertainty degrades perceptual processing. Psychon. Bull. Rev. 14, 522–526, http://dx.doi.org/10.3758/BF03194101.

Rose, G.M., DeVilliers, C., Straub, D.W., 2009. Chronism theory, culture, and system delay: a longitudinal study of post-apartheid South Africa. J. Global Inf. Manag. 17, 1–28, http://dx.doi.org/10.4018/jgim.2009070901.

Schaefer, F., 1990. The effect of system response times on temporal predictability of work flow in human–computer interaction. Hum. Perform. 3, 173–186.

Schleifer, L.M., Amick III, B.C., 1989. System response time and method of pay: stress effects in computer-based tasks. Int. J. Hum. Comput. Interact. 1, 23–39.

Seibold, V.C., Bausenhart, K.M., Rolke, B., Ulrich, R., 2011. Does temporal preparation increase the rate of sensory information accumulation? Acta Psychol. 137, 56–64, http://dx.doi.org/10.1016/j.actpsy.2011.02.006.

Seibold, V.C., Fiedler, A., Rolke, B., 2011. Temporal attention shortens perceptual latency: a temporal prior entry effect. Psychophysiology 48, 708–717, http://dx.doi.org/10.1111/j.1469-8986.2010.01135.x.

Sellier, A.L., Chattopadhyay, A., 2009. Valuing time: moderate download times can improve online goal pursuit. J. Consum. Psychol. 19, 236–245, http://dx.doi.org/10.1016/j.jcps.2009.02.014.

Selvidge, P.R., Chaparro, B.S., Bender, G.T., 2002. The world wide wait: effects of delays on user performance. Int. J. Ind. Ergon. 29, 15–20.

Seneler, C.O., Basoglu, N., Daim, T., 2009. Interface feature prioritization for web services: case of online flight reservations. Comput. Hum. Behav. 25 (4), 862–877, http://dx.doi.org/10.1016/j.chb.2008.12.028.

Seow, S.C., 2008. Designing and Engineering Time: The Psychology of Time Perception in Software. Addison Wesley, Upper Saddle River, NJ.

Shahar, N., Meyer, J., Hildebrandt, M., Rafaely, V., 2012. Detecting system failures from durations and binary cues. Int. J. Hum. Comput. Stud. 70, 552–560.

Simoens, P., Vankeirsbilck, B., Deboosere, L., Ali, F.A., De Turck, F., Dhoedt, B., Demeester, P., 2011. Upstream bandwidth optimization of thin client protocols through latency-aware adaptive user event buffering. Int. J. Commun. Syst. 24, 666–690, http://dx.doi.org/10.1002/dac.1188.

Steinborn, M.B., Langner, R., 2011. Distraction by irrelevant sound during foreperiods selectively impairs temporal preparation. Acta Psychol. 136, 405–418, http://dx.doi.org/10.1016/j.actpsy.2011.01.008.

Steinborn, M.B., Rolke, B., Bratzke, D., Ulrich, R., 2008. Sequential effects within a short foreperiod context: evidence for the conditioning account of temporal preparation. Acta Psychol. 129, 297–307.

Steinborn, M.B., Rolke, B., Bratzke, D., Ulrich, R., 2009. Dynamic adjustment of temporal preparation: shifting warning signal modality attenuates the sequential foreperiod effect. Acta Psychol. 132, 40–47, http://dx.doi.org/10.1016/j.actpsy.2009.06.002.

Steinborn, M.B., Rolke, B., Bratzke, D., Ulrich, R., 2010. The effect of a cross-trial shift of auditory warning signals on the sequential foreperiod effect. Acta Psychol. 134, 94–104.

Steinhauser, M., Kiesel, A., 2011. Performance monitoring and the causal attribution of errors. Cogn. Affect. Behav. Neurosci. 11, 309–320, http://dx.doi.org/10.3758/s13415-011-0033-2.

Szameitat, A.J., Rummel, J., Szameitat, D.P., Sterr, A., 2009. Behavioral and emotional consequences of brief delays in human–computer interaction. Int. J. Hum. Comput. Stud. 67, 561–570.

Thomaschke, R., Dreisbach, G., 2013. Temporal predictability facilitates action, not perception. Psychol. Sci. 24, 1335–1340, http://dx.doi.org/10.1177/0956797612469411.

Thomaschke, R., Kiesel, A., Hoffmann, J., 2011. Response specific temporal expectancy: evidence from a variable foreperiod paradigm. Atten., Percept. Psychophys. 73, 2309–2322, http://dx.doi.org/10.3758/s13414-011-0179-6.

Thomaschke, R., Wagener, A., Kiesel, A., Hoffmann, J., 2011a. The scope and precision of specific temporal expectancy: evidence from a variable foreperiod paradigm. Atten. Percept. Psychophys. 73, 953–964, http://dx.doi.org/10.3758/s13414-010-0079-1.

Thomaschke, R., Wagener, A., Kiesel, A., Hoffmann, J., 2011b. The specificity of temporal expectancy: evidence from a variable foreperiod paradigm. Q. J. Exp. Psychol. 64, 2289–2300, http://dx.doi.org/10.1080/17470218.2011.616212.

Thum, M., Boucsein, W., Kuhmann, W., Ray, W.J., 1995. Standardized task strain and system response times in human–computer interaction. Ergonomics 38, 1342–1351.

Wagener, A., Hoffmann, J., 2010. Temporal cueing of target-identity and target-location. Exp. Psychol. 57, 436–445, http://dx.doi.org/10.1027/1618-3169/a000054.

Watanabe, M., Hirose, K., Den, Y., Minematsu, N., 2008. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. Speech Commun. 50, 81–94.

Weber, F., Haering, C., Thomaschke, R., 2013. Improving the human computer dialogue with increased temporal predictability. Hum. Factors 55, 881–892, http://dx.doi.org/10.1177/0018720813475812.

Wendt, M., Kiesel, A., 2011. Conflict adaptation in time: foreperiods as contextual cues for attentional adjustment. Psychol. Bull. Rev. 18, 910–916.

Whelan, R., 2008. Effective analysis of reaction time data. Psychol. Rec. 58, 475–482.

Williges, R.C., Williges, B.H., 1982. Modeling the human operator in computer-based data entry. Hum. Factors 24, 285–299.

Wundt, W., 1874. Grundzüge der Physiologischen Psychologie. Engelmann, Leipzig.