The effects of previous misestimation of task duration on estimating future task duration

Cornelius J. König
Universität des Saarlandes

Andreja Wirz
Universität Zürich

Kevin E. Thomas
Bournemouth University

Rahel-Zoë Weidmann
Universität Zürich

Cornelius J. König, Universität des Saarlandes, Saarbrücken, Germany; Andreja Wirz and Rahel- Zoë Weidmann, Universität Zürich, Switzerland; Kevin E. Thomas, Bournemouth University, United Kingdom. The first author worked previously at the Universität Zürich. Correspondence concerning this article should be addressed to Cornelius J. König, Universität des Saarlandes, Campus A1 3, 66123 Saarbrücken, Germany, ckoenig@mx.uni-saarland.de, fax: +49-681-302

The effects of previous misestimation of task duration on estimating future task duration

Abstract

It is a common time management problem that people underestimate the duration of tasks, which has been termed the “planning fallacy.” To overcome this, it has been suggested that people should be informed about how long they previously worked on the same task. This study, however, tests whether previous misestimation also affects the duration estimation of a novel task, even if the feedback is only self-generated. To test this, two groups of participants performed two unrelated, laboratory-based tasks in succession. Learning was manipulated by permitting only the experimental group to retrospectively estimate the duration of the first task before predicting the duration of the second task. Results showed that the experimental group underestimated the duration of the second task less than the control group, which indicates a general kind of learning from previous misestimation. The findings imply that people could be trained to carefully observe how much they misestimate task duration in order to stimulate learning. The findings are discussed in relation to the anchoring account of task duration misestimation and the memory-bias account of the planning fallacy.

Keywords: planning fallacy; time management; time estimation

One of the problems that makes time management (Koch and Kleinmann 2002) so difficult is the planning fallacy: the tendency to underestimate future task duration despite knowing that previous tasks overran (Kahneman and Tversky 1979). Considerable research (e.g., Buehler et al. 1997; Burt and Kemp 1994; Halkjelsvik et al. 2011; König 2005; Roy et al. 2008; Thomas et al. 2004; Weick and Guinote 2010; for recent overviews see Buehler et al. 2010, and Halkjelsvik and Jørgensen 2012) has almost universally found that tasks take longer than predicted, and this has been observed on various laboratory and real world tasks including writing college assignments (e.g., Buehler et al. 1994) and shopping for gifts (Kruger and Evans 2004). Such underestimation of task duration may cause serious problems; for example, students may start to work on assignments too late to achieve good grades and gifts bought in the rush may not have the anticipated consequences (see Kruger and Evans 2004).

Kahneman and Tversky (1979) have suggested that two types of data are available when predicting task duration: singular information (i.e., aspects of the focal task) and distributional information (i.e., information about previous tasks). They argue that the planning fallacy occurs because predictions are based on singular information whilst distributional information is ignored or neglected. In support of this account, Buehler et al. (1994) analysed verbal protocols and found that previous tasks were rarely given as reasons for task duration predictions. Moreover, Buehler et al. found that underestimation was reduced when participants were instructed to consider elements common to previous and current/focal tasks. They argue that this technique enabled participants to consider potential impediments to task completion (e.g., computer failure), which resulted in less biased predictions.

An alternative account of the planning fallacy and task duration prediction bias more generally comes from research which indicates that people do consider information about their previous task performance, but still mispredict task duration (König 2005; Thomas and Handley 2008; see also Thomas et al. 2007). This account, which is based on the anchoring and adjustment heuristic (Tversky and Kahneman 1974), posits that predictions are anchored/based on the duration of previous tasks but are insufficiently adjusted according to the demands of the focal task, resulting in the underestimation indicative of the planning fallacy when previous tasks are shorter than focal tasks and overestimation when previous tasks are longer than focal tasks. In support of this account, König (2005) found that prediction bias on a catalogue-searching task differed according to whether a previously-presented number represented a shorter or longer duration, with underestimation or overestimation occurring, respectively. Similarly, Thomas and Handley (2008) found this anchoring effect when numbers represented the duration of similar or different tasks to focal tasks. These studies indicate that information about previous tasks (i.e., duration) can bias predictions of task duration, suggesting that prediction bias is not due to information about previous tasks being neglected or overlooked (Kahneman and Tversky 1979).

The anchoring account (e.g., König 2005) has similarities with the memory-bias account of the planning fallacy (Roy et al. 2005b), which also emphasizes the role of previous task performance or duration. The memory-bias account posits that people mispredict task duration because they base their predictions on incorrect memories about previous, similar tasks. In support of the account, Roy et al. (2008) found that informing people of how long they had previously taken on the same task reduced prediction bias. This suggests that if people are told about the length of previous tasks, their memory is corrected and their predictions on the next task are less biased.

Although the anchoring and memory-bias accounts are valuable additions to the task duration prediction literature by highlighting the role of previous task performance or duration, it seems sensible to question whether the results of the supporting studies generalize to everyday life situations in which people often work on one task and then on another, different task. Participants in experiments may count two stacks of 500 sheets of paper twice (see Roy et al. 2008, Experiment 1) or build two versions of the same toy castle in succession (see Thomas et al. 2007, Experiment 1), but such repetition might not be common in everyday life. For example, even if a researcher has to count the stacks of paper of a paper-and-pencil questionnaire that she wants to distribute, this counting task is likely not to be followed by another counting task for the next study, but by another, different task (e.g., statistical analyses of data for a manuscript).

Furthermore, researchers provided external duration feedback to participants in previous anchoring account studies (Thomas and Handley 2008; see also König 2005) and Roy et al.'s (2008) study, but such external feedback is likely to often be absent from everyday situations. Although telling participants about the duration matched exactly what the aforementioned studies wanted to show, what happens if there is no external feedback? Without external feedback, the duration of a task may be sometimes be salient and sometimes not. If, for example, people regularly consult their watches, they might have a fairly good idea of how long a task took. In other situations, people may only vaguely guess how long they needed to finish a task. This implies that self-generated feedback is likely to be less precise than external feedback.

Nevertheless, self-generated feedback about the duration of a different task might still reduce prediction bias because of people's general capacity to learn from mistakes (e.g., Fischer et al. 2006; Gick and McGarry 1992; Keith and Frese 2008). If they have previously underestimated the duration of a task, they can refer to this insight when making a new prediction about the duration of an upcoming task. For example, a lecturer might realize that it took much longer to grade student essays than expected, and this experience might make the lecturer less optimistic how much time they will need to plan the next lecture. In such a situation, the lecturer learns from previous underestimation and such learning occurs without the intervention of someone else (e.g., a teaching assistant) telling the lecturer how long lecture planning took previously.

That people can learn from previous errors (i.e., via feedback) is a fundamental idea that can be found in many psychological theories. For example, a central idea in Skinner's (1938) learning theory is that positive or negative reinforcement (i.e., feedback) increases or decreases behavior. Thus, if people make an error and get feedback, this decreases the probability of making the error again. Similarly, Carver and Scheier's (1982, 1998) control theory relies on a discrepancy-reducing feedback loop, where feedback about the previous success of behavior is used for decisions about future self-regulated behavior. Our argument is therefore that such learning from errors also happens in the context of making task duration predictions. This learning could be more explicit or more implicit: People might be aware that their last prediction was too optimistic or people might just be sensitized about the danger of making too optimistic predictions, without consciously thinking about it.

The idea that people learn from previous underestimation goes beyond the anchoring (e.g., Thomas and Handley 2008) and memory-bias accounts (Roy et al. 2005b), which require the presence of some information upon which the prediction of the upcoming task is based (i.e., a previously-presented duration anchor or misremembered memory trace of a previous task). The accounts therefore imply that underestimation will not occur or prediction bias will be reduced when there is no previous task experience. However, as anchor values that are self-generated can influence non-temporal judgements (e.g., Epley and Gilovich 2005), anchoring could also explain why learning from mistakes on previous tasks can influence task duration prediction bias. This would happen by generating time predictions based on the perceived duration of previous tasks and adjusting these predictions according to the demands of the focal task. This self-learning effect could reduce or increase prediction bias depending on the relative durations of previous and focal tasks but would occur regardless of the similarity of the focal task to previous tasks.

To test this self-learning effect, we had participants work on two dissimilar tasks. All participants had to estimate prospectively the duration of the first task, but only half of them were also asked to retrospectively estimate the duration. Then, we measured how much participants underestimated or overestimated the new, upcoming task. More formally, we hypothesised:

> *The duration of an upcoming new task is misestimated less if participants have to retrospectively estimate the duration of a just-completed task than if they do not have to retrospectively estimate the duration of a just-completed task.*

Method

Participants

The sample consisted of 64 students from a Swiss university, who participated voluntarily. One participant had to be excluded from the data analyses according to an outlier analysis (i.e., a large Cook-distance value). Of the remaining 63 participants, 53 were female and 10 were male. The sample's average age was 24.59 years (*SD* = 6.41). Most of the participants were psychology students (*n* = 57), three were studying journalism, one studied agronomy, one political science, and one law. Forty-three participants had just started studying, 10 were in the third and 10 in the fifth or in a higher semester of their studies. All participants were entered in a prize draw to win a watch, and psychology students also received course credit for participating.

Tasks

Two tasks were used: a LEGO® assembly task (with LEGO® being a trademark of the LEGO Group, Billund, Denmark) and a colouring task. The two tasks were counterbalanced. Task selection was based on the following three criteria. First, as has been shown (e.g. Roy et al. 2005b), the tendency to underestimate the duration of a task can be observed when participants need at least 12.5 minutes to complete a task. Hence, the tasks used in this study should last at least 12.5 minutes. Second, so that participants could not rely on the duration of the just-completed first task when predicting how long it would take them to complete the second task, the tasks should be dissimilar. Third, it might be problematic if task duration can be calculated on the basis of countable units. As we will outline below, both tasks fulfilled these criteria.

LEGO® assembly task. Following Hinds (1999) and Roy (2003), participants had to piece together the 81-piece bird of prey figure from the LEGO® Creator Set “Wild animals” (No. 4884-1). Roughly a third of all pieces in the set are needed to assemble this animal. In a pretest (N = 19), participants needed 25.32 minutes on average (SD = 6.75) to complete this task. Furthermore, they underestimated (in minutes) the task duration both prospectively (M = 16.74, SD = 13.54) and retrospectively (M = 22.63, SD = 6.27). Hence, it can be assumed that there is not a tendency to overestimate the duration of this task.

In the present study, the experimenter put the whole LEGO® Creator Set “Wild animals” on a table. Then, the experimenter classified pieces in the colours light grey, dark grey and black according to a picture of the completed bird of prey in order to avoid confusion regarding these colours. After that, the experimenter showed the participant the picture of the completed bird of prey and asked the participant to estimate how many minutes it would take them to assemble that LEGO® figure with the aid of an instruction manual. Afterwards, the participant was asked to start with the task and to complete it properly.

Colouring task. This involved colouring a dinosaur template according to a coloured master copy. In a pre-test (*N* = 8), the average time for completion of this task was 22.63 minutes (*SD* = 5.18), which exceeds the necessary 12.5 minutes (Roy et al. 2005b). In the study, the experimenter first presented the coloured master copy as well as the to-be-coloured template to participants and prepared the coloured crayons and the pencil sharpener. Again, the experimenter compared the yellow, orange and red coloured crayons, which were easy to confuse, with the corresponding colours on the master copy. After that, participants were asked (in everyday language) to colour the template exactly according to the master copy, but neither to be too exacting nor to make mistakes.

Procedure

Data were collected in individual sessions for which one and a half hours were scheduled. Participants were informed that the study was about time management and that it addressed the question of how people deal with time. Participants would complete different tasks and fill in questionnaires. In order to avoid participants being able to compute the duration of the two tasks based on the duration of the whole experiment, they were not informed about how many tasks they would have to complete and what exactly was measured.

At the beginning of the experimental session, participants were asked to switch off their cellular phone and to put their watch out of sight to reduce interferences during the data collection. Furthermore, they were asked whether they suffered from colour blindness. One participant admitted colour-blindness but identified all colours correctly when tested. Since the data of this participant were not peculiar in any way, this participant was not excluded from the data analyses.

Participants were randomly assigned to four groups: two groups (an experimental and a control group) starting with the LEGO® task and two groups (an experimental and a control group) starting with the colouring task (see also Figure 1 for a flow chart for the main part of the study). First, all participants were asked to predict how many minutes it would take them to complete the first task (LEGO® or colouring task) and then to perform it. Only the experimental groups also had to estimate the duration of the first task retrospectively. Immediately after this, all groups were introduced to the second task (task not performed first) and asked to predict its duration and then to perform it. The actual times for completion of the first and second tasks were unobtrusively measured by the experimenter using a stopwatch. When the participants signalled completion of the task, the stopwatch was stopped and the experimenter checked whether the task was properly completed. If a task was not finished appropriately (e.g., pieces of the bird were still missing), participants were asked to correct such errors and the stopwatch was re-started from the point at which it had been stopped. All participants were also asked to estimate the duration of the second task retrospectively.

After the second task, control question was asked, participants had to indicate on a five-point Likert-scale how often they used to play with LEGO® bricks and how often they used to draw, respectively, in their childhood and how often they had done these activities since their childhood (1 = never, 5 = very often). All participants indicated that they used to draw more often (M = 3.97, SD = 1.12) than they used to play with LEGO® (M = 3.17, SD = 1.21) as a child, $t(62) = -3.78$, $p < .001$. Participants also indicated more drawing (M = 2.98, SD = 1.09) than LEGO® playing (M = 1.49, SD = .78) since childhood, $t(62) = -9.16$, $p < .001$. More importantly, however, the two groups did not differ in their childhood experience of playing with LEGO® or with drawing, both *t*s < 1, and they did not differ in the extent of LEGO® playing or drawing in later years, both *t*s < 1.

Since we had invited participants for a study of one and a half hours and since we did not want participants to use this information when estimating task duration, we filled the remaining time (if necessary) with a "two minutes task" and a questionnaire. The "two minute task" means that participants had to state when a two-minute period was over (see Block and Zakay 1997). The experimenter indicated the start of the two minutes and, at the same time, activated a stopwatch. When participants signalled the assumed end of the two minutes, the experimenter stopped the stopwatch. Furthermore, participants were kept busy by filling out the Time Management Behavior Scale (Macan 1994, in its German version, König and Kleinmann 2006).

Dependent Variable

To measure misestimation, we used the proportional error (Roy 2003), which is calculated as follows:

$$\text{Proportional error of predicted task duration } (PE_{prediction}) = \frac{\text{predicted task duration} - \text{actual task duration}}{\text{actual task duration}}$$

A negative proportional error indicates underestimation of task duration and a positive proportional error indicates overestimation. If the proportional error is zero, then estimated and actual task duration are equal (i.e., no bias). Thus, the greater the proportional error deviates from zero, the larger the bias.

importantly, however, the two groups did not differ in their childhood experience of playing with

Table 1 contains basic descriptive statistics. The negative values of the two proportional error of predicted duration in Table 1 show that participants underestimated the duration of the both tasked prospectively. We tested our hypothesis via a 2 (task 1 vs. task 2) × 2 (experimental vs. control group) × 2 (order of tasks) mixed-model ANOVA, with the proportional error of predicted task duration as the dependent variable. Order of tasks is a control variable that captured whether participants started with the LEGO® task or the colouring task. Task 1 vs. task 2 is the within-subject variable and experimental vs. control group and order of tasks the between-subject variables. As Table 2 shows, our hypothesis was supported: there was a significant task 1 vs. task 2 × experimental vs. control group interaction. The estimated marginal means of this interaction were -0.36 for task 1 in the control group, -0.28 for task 2 in the control group, -0.42 for task 1 in the experimental group, and -0.15 for task 2 in the experimental group, and a post-hoc comparison with a Sidak correction revealed that only the difference within the experimental group was significant ($p < .01$). This indicates that if participants had to estimate the duration of a just-completed task, they underestimated the duration of a new and different task less than if they did not have to estimate the duration of a just-completed task.

Furthermore, the ANOVA showed a main effect of task 1 vs. task 2 (i.e., there was an general decrease in the proportional error of predicted task duration from task 1 to task 2), but no main effects for experimental vs. control group variable nor for the control variable (i.e., order or tasks). In addition, the task 1 vs. task 2 × order of tasks interaction was significant. The estimated marginal means of this interaction were -0.44 for the LEGO® task as the first task, -0.33 for the colouring task as the first task, -0.16 for colouring task as the second task, and -0.28 for the LEGO® task as the second task, and post-hoc comparison with a Sidak correction revealed that only the within-group difference between the LEGO® task as the first task and the colouring task as the second task was significant ($p < .01$). This finding suggests that it was those who underestimated the LEGO® task in the beginning who learned from previous misestimation.

Discussion

This study tested the hypothesis that people can learn from misestimating the duration of previous, dissimilar tasks. There was indeed evidence for this: Participants underestimated the duration of a new and different upcoming task to a lesser extent when they had retrospectively estimated the duration of a just-completed task than if they did not make a retrospective estimate. This implies that people realise that a previous task took longer than expected and use this information when predicting the duration of a new, unrelated task. Thus, our results suggest that people can learn from previous misestimation.

The reduction in prediction bias on the second task when the duration of the first task was retrospectively estimated is consistent with our claim that self-generated anchor values can influence task duration predictions as well as non-temporal judgements (Epley and Gilovich 2005). Specifically, the act of generating a duration estimate for the just-completed, first task led participants in the experimental group to give longer and less biased predictions on the upcoming second task. Anchoring the retrospective estimate on the perceived duration of the first task and using this value as a basis for the duration prediction on the second task would be expected to result in less prediction bias in the experimental group because retrospective estimates on the first task were close to actual durations of the first task and the durations of the first and second tasks were similar. Thus, it seems that anchoring predictions on the perceived duration of previous tasks can explain why learning from mistakes on previous tasks influences task duration prediction bias. Moreover, using perceived task duration (rather than other task information; see Thomas et al. 2007) as an anchor means that this anchoring effect occurs regardless of the similarity of previous and upcoming tasks. Importantly, our results suggest that the anchoring account (e.g., Thomas and Handley 2008) applies to task duration prediction situations in which there is no explicit, externally-presented duration information or feedback.

Despite being a minimal manipulation (Cortina and Landis 2009; Prentice and Miller 1992), it was effective, with the only difference between the groups being the retrospective estimation on the first task in the experimental group – an intervention of approximately 10 seconds. Participants were never asked to contemplate these estimates, and the experiment continued immediately after these estimates were made. However, the mere act of having to provide an estimate of previous task duration was found to be sufficient to reduce prediction bias on a very different task performed subsequently.

Our study can also be considered a conservative test of our hypothesis because participants in the control group may (or may not) have had thoughts about how long they had taken to finish the first task, the duration of which they had to estimate before performing that task. Thus, these participants may have also wondered whether the first task took them longer than predicted. If this was the case, they could also interpret their own bewilderment about being slower than expected as a sign that they should be less optimistic when making a duration prediction on the second task. Indeed, we found that prediction bias was less on the second task than the first task regardless of the retrospective estimation manipulation and which task was performed first, which supports our proposed self-learning effect. This finding implies that participants in the control and experimental groups took account of the perceived duration of the first task when predicting the duration of the second task, but that this self-learning effect was enhanced by the act of generating a retrospective estimate on the first task. However, the fact that the control group underestimated the duration of the first and second tasks to a similar extent indicates that people may not spontaneously use retrospective estimates. By conducting verbal protocol analyses to ascertain the reasons given for task duration estimates, future research will be well-placed to clarify this aspect of our results.

In addition, our results suggests factors beyond those suggested by the memory-bias account (Roy et al. 2005b). According to Roy et al, it is the misremembered memory trace of a previous task that leads to the underestimation of a new, similar task; but if the length of a new task is estimated, there cannot be a misleading memory (see also Roy and Christenfeld 2007). Our study employed two dissimilar tasks, and the memory-bias account would thus not predict an effect of our manipulation. This implies that the memory-bias account can only explain some but not all effects of the planning fallacy phenomenon and task duration misestimation. (We hasten to add such a position is consistent with proponents of the memory bias account, see, e.g., Roy et al. 2005a).

Given our findings, a number of avenues exist for future research. In particular, researchers could examine the mechanism underpinning learning from previous misestimation. Although our results show that such learning occurs, it remains debatable how explicit (or implicit) this learning was - whether participants were merely sensitized to be more cautious when making a prediction on the second task or whether participants consciously decided to be more cautious in that prediction. Future research could begin to examine the underlying mechanism by doing things such as analysing participants’ reported thoughts and manipulating whether or not participants predict task duration under time pressure.

Future research could also seek to identify the conditions under which learning from previous misestimation is more or less likely to occur. In particular, similarity between tasks may be one such moderating variable because people may be reluctant to transfer the insight of having misestimated the duration of a task to another task if they perceive tasks as dissimilar. For instance, students may not transfer misestimation experiences from home (e.g., preparing dinner) to university (e.g., preparing for a course) because they may consider preparing dinner and preparing for a course as belonging to two separates spheres and thus too dissimilar for the transfer of this insight to be useful. People may also be less likely to transfer if one task is much longer than the other (e.g., students writing a 1000-word essay versus writing a 10000-word dissertation) because they are perceived as being too dissimilar (although the same psychological processes may cause misestimation, e.g., Kahneman and Tversky 1979; Roy et al. 2005b). Another moderating variable may the temporal distance between tasks (cf. Roy and Christenfeld 2007). In our study, participants estimated the duration of the second task directly after finishing the first task, which made learning particularly likely. However, learning may become less likely the longer the inter-task temporal interval.

Another exciting avenue for research would be to study tasks of very short duration, which are known to be typically overestimated (Rodon and Meyer 2012; Roy et al. 2005b; cf. Halkjelsvik and Jørgensen 2012). Although studying underestimation is more relevant from a practical point of view (e.g., customers are likely only to complain if the renovation of a store is finished several days after rather than before the announced opening day), the psychological process that became evident in our study (i.e., that people can learn from experience) should also apply to overestimation. For example, the duration of a task that takes only three minutes is likely to be overestimated, and if people learn from this misestimation, they should make shorter predictions about how long a second subsequent task takes.

Furthermore, if people can learn from experience of estimating in general, as our research shows, this suggests that experience with estimating the duration of a variety of tasks might matter. One could even speculate that older people might be less prone to the misestimation of task duration because of their life experience with a variety of tasks - that becoming wiser over the years might also include less biased predictions of task duration. To our knowledge, age effects have not been studied in the planning fallacy literature, but this could another interesting direction for future research.

We also like to call for more field research in the field of duration misestimation. Laboratory studies like this one are always limited because they study phenomena in artificial environments and often manipulate only one variable. Although the latter can also be seen as the strength of the laboratory experiment, we hope that future research finds why to study this in the field, ideally taking many influential variables into account.

Our study also has an important applied implication because its findings could be used in time management training courses. Within the time management literature (e.g., Claessens et al. 2007; Koch and Kleinmann 2002), the importance of good planning has been acknowledged as a major issue. Our study suggests that time management course attendees could be trained to carefully observe how biased their estimates of task duration are in order to stimulate learning from these experiences. Furthermore, people should be encouraged to develop a routine of retrospective monitoring of how much they underestimated previous task duration so that they can keep such information in mind when estimating the duration of upcoming tasks. Given that our manipulation took less than 10 seconds but was effective, such monitoring should be easily incorporated into daily life.

Our study provides further evidence of the pivotal role of previous task performance in the task duration prediction process as well as supporting and extending the anchoring account of misestimation (e.g., Thomas and Handley 2008). Our results suggest that prediction bias is reduced when people generate feedback on the perceived duration of a previous task and use this performance information as an anchor for estimating the duration of san upcoming task. Given the cost of underestimating the duration of major projects, the need to more fully understand the task duration prediction process cannot be overstated.

References

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin and Review, 4*, 184-197, doi:10.3758/BF03209393.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin*, *23*, 238-247, doi:10.1177/0146167297233003.

Buehler, R., Griffin, D., & Peetz, J. (2010). The planning fallacy: Cognitive, motivational, and social origins. In P. Z. Mark, & M. O. James (Eds.), *Advances in experimental social psychology* (pp. 1-62, Vol. 43). New York, NY: Academic Press.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67*, 366-381, doi:10.1037/0022-3514.67.3.366.

Burt, C. D. B., & Kemp, S. (1994). Construction of activity duration and time management potential. *Applied Cognitive Psychology, 8*, 155-168, doi:10.1002/acp.2350080206.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychological Bulletin*, *92*, 111-135, doi:10.1037/0033-2909.92.1.111.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York, NY: Cambridge University Press.

Claessens, B. J. C., van Eerde, W., & Rutte, C. G. (2007). A review of the time management literature. *Personnel Review*, *36*, 255-276, doi:10.1108/00483480710726136.

Cortina, J. M., & Landis, R. S. (2009). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance, & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 287-308). New York, NY: Routledge.

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making, 18*, 199-212, doi:10.1002/bdm.495.

Fischer, M. A., Mazor, K. M., Baril, J., Alper, E., DeMarco, D., & Pugnaire, M. (2006). Learning from mistakes: Factors that influence how students and residents learn from medical errors. *Journal of General Internal Medicine, 21*, 419-423, doi:10.1111/j.1525-1497.2006.00420.x.

Gick, M. L., & McGarry, S. J. (1992). Learning from mistakes: Inducing analogous solution failures to a source problem produces later successes in analogical transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 623-639, doi:10.1037/0278-7393.18.3.623.

Halkjelsvik, T., & Jørgensen, M. (2012). From origami to software development: A review of studies on judgment-based predictions of performance time. *Psychological Bulletin, 138*, 238-271, doi:10.1037/a0025996.

Halkjelsvik, T., Jørgensen, M., & Teigen, K. H. (2011). To read two pages, I need 5 minutes, but give me 5 minutes and I will read four: How to change productivity estimates by inverting the question. *Applied Cognitive Psychology, 25*, 314-323, doi:10.1002/acp.1693.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, *5*, 205-221, doi:10.1037/1076-898X.5.2.205.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in the Management Sciences*, *12*, 313-327.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*, 59-69, doi:10.1037/0021-9010.93.1.59.

Koch, C. J., & Kleinmann, M. (2002). A stitch in time saves nine: Behavioural decision-making explanations for time management problems. *European Journal of Work and Organizational Psychology*, *11*, 199-217, doi:10.1080/13594320244000120.

König, C. J. (2005). Anchors distort estimates of expected duration. *Psychological Reports*, *96*, 253-256, doi:10.2466/PR0.96.2.253-256.

König, C. J., & Kleinmann, M. (2006). Individual differences in the use of time management mechanics and in time discounting. *Individual Differences Research*, *4*, 194-207.

Kruger, J., & Evans, M. (2004). If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *Journal of Experimental Social Psychology*, *40*, 586-598, doi:10.1016/j.jesp.2003.11.001.

Macan, T. H. (1994). Time management: Test of a process model. *Journal of Applied Psychology*, *79*, 381-391, doi:10.1037/0021-9010.79.3.381.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*, 160-164, doi:10.1037/0033-2909.112.1.160.

Rodon, C., & Meyer, T. (2012). Searching information on the web and planning fallacy: A pilot investigation of pessimistic forecasts. *European Review of Applied Psychology*, *62*, 103-109, doi:10.1016/j.erap.2011.12.004.

Roy, M. M. (2003). *Memory bias: Why we underestimate the duration of future events*. PhD thesis, University of California, San Diego,

Roy, M. M., & Christenfeld, N. J. S. (2007). Bias in memory predicts bias in estimation of future task duration. *Memory & Cognition*, *35*, 557-564, doi:10.3758/BF03193294.

Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005a). The broad applicability of memory bias and its coexistence with the planning fallacy: Reply to Griffin and Buehler (2005). *Psychological Bulletin, 131*, 761-762, doi:10.1037/0033-2909.131.5.761.

Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005b). Underestimating the duration of future events: Memory incorrectly used or memory bias? *Psychological Bulletin, 131*, 738-756, doi:10.1037/0033-2909.131.5.738.

Roy, M. M., Mitten, S. T., & Christenfeld, N. J. S. (2008). Correcting memory improves accuracy of predicted task duration. *Journal of Experimental Psychology: Applied, 14,* 266-275, doi:10.1037/1076-898X.14.3.266.

Skinner, B. F. (1938). *The behavior of organisms*. New York, NY: Appleton-Century-Crofts.

Thomas, K. E., & Handley, S. J. (2008). Anchoring in time estimation. *Acta Psychologica, 127*, 24-29, doi:10.1016/j.actpsy.2006.12.004.

Thomas, K. E., Handley, S. J., & Newstead, S. E. (2004). The effects of prior experience on estimating the duration of simple tasks. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 22*, 83-100.

Thomas, K. E., Handley, S. J., & Newstead, S. E. (2007). The role of prior task experience in temporal misestimation. *Quarterly Journal of Experimental Psychology*, *60*, 230-240, doi:10.1080/17470210600785091.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131, doi:10.1126/science.185.4157.1124.

Weick, M., & Guinote, A. (2010). How long will it take? Power biases time predictions. *Journal of Experimental Social Psychology*, *46*, 595-604, doi:10.1016/j.jesp.2010.03.005.

Table 1

Means and Standard Deviations of actual, predicted and recalled duration (in minutes) and predicted and recalled proportional errors per group and task.

	Control group (*n* = 31)		Experimental group (*n* = 32)	
	M	*SD*	*M*	*SD*
Task 1				
Predicted duration	15.58	7.44	14.59	9.59
Actual duration	25.15	8.81	26.58	11.60
Retrospective duration	-	-	24.00	12.61
$PE_{prediction}$	-0.36	0.25	-0.43	0.24
$PE_{retrospection}$[a]	-	-	-0.09	0.27
Task 2				
Predicted duration	17.87	7.78	20.34	6.75
Actual duration	25.23	8.32	26.23	11.22
Retrospective duration	26.29	12.15	27.03	9.68
$PE_{prediction}$	-0.28	0.27	-0.15	0.32
$PE_{retrospection}$[a]	0.03	0.27	0.09	0.35

Note. PE = proportional error. A negative proportional error indicates underestimation of the task duration, a positive value overestimation. For half of the participants, task 1 was the LEGO® assembly task; whereas task 1 was the colouring task for the other half.

[a] The proportional error of retrospective task duration ($PE_{retrospection}$) was calculated analogously as the proportional error of predicted task duration.

Table 2

Results of a Mixed-Model ANOVA with the Proportional Error of Predicted Task Duration as Dependent Variables

Source	*df*	*SS*	*MS*	*F*	$\eta^2_{partial}$
	Within subjects				
Task 1 vs. task 2	1	0.90	0.90	16.07**	.21
Task 1 vs. task 2 × Experimental vs. control group	1	0.28	0.28	5.04*	.08
Task 1 vs. task 2 × Order of tasks	1	0.42	0.42	7.45**	.11
Task 1 vs. task 2 × Experimental vs. control group × Order of tasks	1	0.11	0.11	1.95 n.s.	.03
Error 1	59	3.31	0.06		
	Between subjects				
Experimental vs. control group	1	0.04	0.04	0.43 n.s.	.01
Order of tasks	1	0.00	0.00	0.10 n.s.	.00
Experimental vs. control group × Order of tasks	1	0.07	0.07	0.83 n.s.	.01
Error 2	59	5.08	0.09		

Note: $N = 63$. Order of tasks, a control variable, captured whether participants started with the LEGO® task or the colouring task.

* $p < .05$; ** $p < .01$.

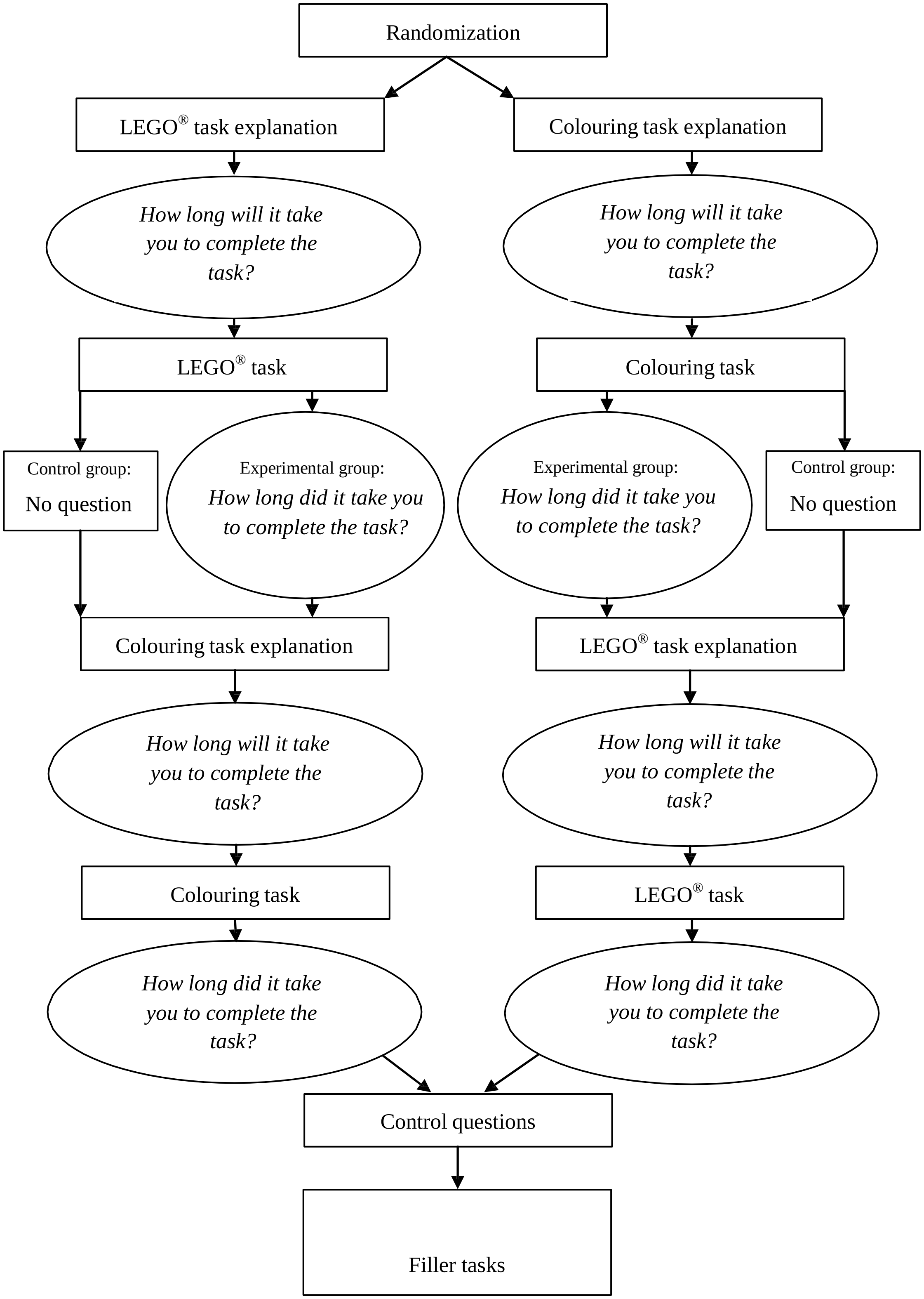

Figure 1. Flowchart of the experiment.

3628.