**Supplementary Materials (SM)**

for

**The effect of time ambiguity on choice depends on delay and amount magnitude**

Iris Ikink1,2,3, Karin Roelofs1,2, Bernd Figner1,2

1 Behavioural Science Institute, Radboud University Nijmegen, The Netherlands

2 Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

3Ghent University, Ghent, Belgium

Overview of SM content:

|  |  |
| --- | --- |
| Appendix A | More details on hypothesis 5 (H5)  |
| Appendix B | Computational modeling – methods and results |
| Table A | Overview of model fits for all tested computational models |
| Table B | Comparison of statistical model fit across the 3 ways of time-ambiguity encoding using the As-Reported-In-Manuscript (ARIM) sample versus either (1) all trials or (2) all participants  |
| Table C | Comparison of results of the main choice model using the ARIM sample versus either (1) all trials or (2) all participants  |
| Table D | Comparison of results of the Possible Today Delivery (PTD) model using the ARIM sample versus all participants |
| Table E | Overview of the 39 unique trials (given delay midpoint and time-ambiguity level) |
| Figure A | Display of the one odd timeline trial (excluded in the ARIM sample) |
| Table F | Results follow-up model for the interaction between time-ambiguity and delay |
| Table G | Results follow-up model for the interaction between time-ambiguity and amount |
| Table H | Results follow-up model for the 3-way interaction between time-ambiguity, delay, and amount |
| Table I | Results follow-up model for interaction between display and amount |
| Table J | Results follow-up model for interaction between display and delay |
| Appendix C | Details of the interaction effects between display and time-ambiguity |
| Table K | Results follow-up model for the 3-way interaction between display, time-ambiguity, and delay |
| Table L | Results follow-up model for the 4-way interaction between display, time-ambiguity, delay, and amount |
| Table M  | Proportion of LL choices per delay midpoint, time-ambiguity level, amount, and display version in the combined sample (*N*=194)  |
| Appendix D | Details of follow-up models for the possible today-delivery model |
| Table N | Overview of the included trials in Restricted Dataset 2, to test the Possible-Today Delivery (PTD) hypothesis in an alternative way (i.e., follow-up model 1) |
| Appendix E | Overview of the 7 items used in Study 3  |
| Table O | Descriptive statistics of the 7 items used in Study 3 (*N*=202)  |

**Appendix A: More details on hypothesis 5 (H5)**

In this section, we explain in more detail what our hypotheses for the effects of display type are based on. The main reason to include two display versions was based on results of the study of Ikink, Engelmann, van den Bos, Roelofs, and Figner (2019). In that study, participants first completed two pre-scanner tasks to establish indifference pairs (based on exact delays), which were then used in the fMRI task with and without time-ambiguous intertemporal choice trials. Interestingly, in the fMRI task participants chose more patiently in the exact intertemporal choice trials than was expected given the estimated indifference pairs. We hypothesized that this increased patience during the fMRI task compared to the pre-scanner tasks could perhaps be caused by a difference in how the trials were presented: The fMRI task had a visuospatial (i.e., timeline) representation format in order to visualize the time-ambiguity (inspired by Tymula et al., 2012), while the pre-scanner tasks used the more typical verbal-numerical (i.e., word) representation format.

An increased level of patience in a timeline compared to a word display could be due to several reasons. First, timelines may lead to an anchoring bias. I.e., the maximum delay could become the reference point against which all other (shorter) delays are evaluated, which might therefore appear to be relatively short (Furnham & Boo, 2011). After all, the maximum delay is always visible when a timeline display is used, whereas this is not the case in a word display (even though all participants were reminded before the task started what the maximum delay in the task was). Secondly, the visual representation of an SS and/or LL option on a timeline may make delayed options more concrete, which in turn could lead to more patient choices. For example, in the date/delay effect, describing delays as more concrete calendar dates instead of their less concrete duration in e.g., days or months also leads to increased patience (Read, Frederick, Orsel, & Rahman, 2005). The increase in concreteness may lead to perceiving a calendar date-delay (or, in our case, a delay as visualized by a timeline) as more near-future and/or result in more episodic future thinking, in turn causing more patient choices (Peters & Büchel, 2009). Lastly, it could be that relatively more attention is given to amount than delay (and/or time-ambiguity) information in a timeline compared to a word display, because in the timeline display the amounts are in somewhat larger font, presented separately from the timeline, and do not need interpretation to understand – unlike the delays and time-ambiguity levels in the timelines (e.g., p. 237, Figure 1 in Ikink et al., 2019 and Fig. 1 in the main text of the current paper). Thus, amounts may stand out more in a timeline display compared to a word display, and as such receive more attention, potentially leading to more patient choices.

Thus, to test the hypothesis that a timeline display would overall lead to more LL choices compared to using a word display, we included two display versions of the task: one using a verbal-numerical display (which we call the word version), the other using a visuospatial display (which we call the timeline version; see also Fig. 1). Furthermore, in Ikink et al. (2019) only timelines were used in the intertemporal choice task including time-ambiguity, but we thought it would also be interesting to investigate whether *time-ambiguity* effects might also differ between a word versus timeline display. After all, displaying time-ambiguity on a timeline might not only make the delays, but also the time-ambiguity ranges more salient and/or concrete, thereby possibly increasing the hypothesized time-ambiguity effects.

**Appendix B: Computational modeling – methods and results**

In an effort to better understand how time-ambiguity might influence subjective value computations, we tested in total 38 different computational choice models. First, we tested all the models previously developed and tested by Ikink et al. (2019) to see if we could replicate their results. As none of these models performed well (i.e., none of these models was able to capture the novel effects observed in the current study), we then explored a small number of new models that at least in principle could have been able to account for the observed time-ambiguity effects in our data. Here, we explain how these models were built up. All their respective formulas can be found in SM-Table A.

To replicate the computational choice models of Ikink et al. (2019), we followed their approach by building on the standard hyperbolic discounting model (Mazur, 1987):

$$SV=\frac{OV}{1+k\*D} \left(Eq.1\right)$$

where *SV* is the subjective value, *OV* is the objective monetary amount, *k* is thediscount rate, and *D* is the delay. The parameter *k* quantifies the level of discounting, with *k*=0 indicating no delay discounting and higher values indicating steeper discounting (i.e., impatience).

Furthermore, unlike Ikink et al. (2019), we also added a probabilistic discounting model (Green, Myerson, & Ostaszewski, 1999) as base model; this model has an additional parameter *s* to account for nonlinear scaling of amount and delay. Given that we found interactions between time-ambiguity level and amount and delay, we thought this model might provide a good alternative model to build on:

$$SV=\frac{OV}{(1+k\*D)^{s}} \left(Eq.2\right)$$

If *s* = 1, equations 1 and 2 are identical; while typically, *s* ≤ 1 (Green et al., 1999).

Using the *mle2*-function from the R-package stats4 (R Core Team, 2018), we fitted the data per participant by estimating the probability of choosing the LL option (PrLLchoice). We computed PrLLchoice using a logistic choice rule as a function of the difference in SV between the LL (SVLL) and SS (SVSS). Furthermore, we included a noise parameter θ to account for choice stochasticity/noisiness in participants’ choices:

$$Pr (LLchoice) = \frac{1}{1+ e^{-θ\*\left(SVLL-SVss\right)}} (Eq.3)$$

Higher values of θ indicate greater consistency (i.e., less noise) in participants’ choices.

Ikink et al. (2019) generated different computational models that describe how time-ambiguity preferences might influence choice via the value function (Eq. 1 – and in this paper as alternative value function also Eq. 2), the choice function (Eq. 3), or both. Likewise, we tested three main types of models in the current paper: Time-ambiguity influencing SV via (i) influencing how the time of delivery is estimated (e.g., being optimistic or pessimistic about the actual delivery time when confronted with a time-ambiguity range, and therefore perceiving the delay to be shorter or longer than its midpoint), (ii) an additive time-ambiguity bonus or penalty that is added to the hyperbolically discounted subjective value (e.g., based on either presence/absence or the extent of time-ambiguity, a time-ambiguous option might be valued as higher or lower), or (iii) time-ambiguity influencing the noise term in the choice rule (e.g., participants might become less consistent in their choice preferences when time-ambiguity is present). To see if we could replicate the results of Ikink et al., we also included two models where both the value function *and* choice function included an additional time-ambiguity parameter, because those two models provided the best fit in their study. We excluded 4 of Ikink et al.’s original models, because in the current study time-ambiguity was only present on the LL option, and therefore models estimating separate parameters for time-ambiguity on the SS and LL could not be included. This resulted in testing 24 computational models in total. All these models predicted slightly different time-ambiguity effects, such as weaker, stronger, or stable/non-changing time-ambiguity effects given longer delays, and/or impacted choice stochasticity.

For models that included time-ambiguity, βSV was the parameter that accounts for a person’s time-ambiguity preference (see SM-Table A for all the formulas). In the preregistered models, each model variant always had four versions; one based on adjusting the formula in Eq. 1, and one based on adjusting the formula in Eq. 2 (i.e., adding parameter *s* or not). Furthermore, we tested for each of those models whether the impact of time-ambiguity on subjective value was better modeled by a discrete relationship (i.e., presence versus absence of time-ambiguity, using if/else coding), or a dose-relationship (i.e., the extent of time-ambiguity in days). In the latter type of model, variable *A* always indicated the respective time-ambiguity level in days. In most (but not all) models, βSV = 0 indicates time-ambiguity neutrality, βSV > 0 indicates time-ambiguity aversion, and βSV < 0 indicates time-ambiguity seeking. The parameter βnoise reflects changes in the stochasticity of participants’ choices under time-ambiguity, via interacting with θ. βnoise = 1 indicates neutrality (i.e., not affecting θ), whereas βnoise < 1 indicates that participants become less consistent (i.e., noisier) when time-ambiguity is present, and βnoise > 1 indicates that participants become more consistent in their choices under time-ambiguity.

After obtaining the result that none of these time-ambiguity models fitted the data better than the standard discounting models (based on BIC comparisons; see SM-Table A), we explored 14 other time-ambiguity models that were informed by the behavioral results that we had found so far in the statistical models (we created an amendment to our preregistration to describe this; <https://osf.io/az95y/>). For example, in the statistical mixed-effects choice models, we did not find a main effect of time-ambiguity aversion, but instead an interaction between time-ambiguity and delay, such that more time-ambiguity was disliked across shorter delays and became liked at longer delays (i.e., a crossover effect). Furthermore, we observed that when the time-ambiguity range included a possible today-delivery, there was little to no effect of time-ambiguity, whereas if the time-ambiguity range did not include a possible today-delivery, the aversive effect of time-ambiguity was stronger. None of our previous time-ambiguity models could capture these effects, possibly explaining why those time-ambiguity models did not outperform the standard discounting models.

Therefore, we tested two new classes of (data-driven, explorative) models: the crossover model and the possible today-delivery (PTD) model. In the crossover model, we adjusted the standard exponential discounting model (Samuelson, 1937):

$$SV=OV\* e^{-\left(k\*D\right)} (Eq.4)$$

Exponential models predict that for each unit increase in the delay there is a fixed percentage decrease in the subjective value of that reward (McKerchar & Renda, 2012). Thus, the difference in SV for waiting 5 vs. 10 days is equally big as the difference in SV for waiting 1000 vs. 1005 days. Previous research has shown that people often discount rewards in a hyperbolic fashion, such that the subjective value of a delayed reward declines more rapidly across shorter delays than across longer delays. However, by using this exponential discounting model we could create the crossover effect we are interested in. This specific model variant is based on a model from Ebert and Prelec (2007):

$$SV=OV\* e^{-(a\*D)^{b}} (Eq.5)$$

where *a* captures the level of impatience (like *k*), and *b* captures time sensitivity. Peters, Miedl, and Büchel (2012) compared different model variants against each other in a large sample of time-exact delay discounting datasets (including the hyperbolic and exponential formulas from equations 1, 2, 4, and 5). They found that the model by Ebert and Prelec provided the best fit, and by extending this model to include time-ambiguity, we would (at least in principle) be able to capture the crossover effect that we observed in our data.

We used two different versions of the crossover model, where for version 1 the formula is the following:

$$SV=OV\* e^{-(k\*D)^{(1+βsv\*\frac{A}{D})}} (Eq.6)$$

Here, βSV is the time-ambiguity parameter, and A the amount of time-ambiguity. Unlike previous models, here we use βSV \* A/D instead of βSV \* A, because we used simulations to explore beforehand how variations in βSV, A, and *k* would impact SV, and the crossover effect turned out to be too extreme when using βSV \* A.

In version 2, we used an if/else formulation to account for the presence versus absence of time-ambiguity. When no time-ambiguity was present (time-exact trials), we used the formula in equation 4, and if time-ambiguity was present (time-ambiguous trials), we used this equation:

$$SV=OV\* e^{-(k\*D)^{(1+βsv)}} (Eq.7)$$

Second, to account for the possible today-delivery effect, we used two new variants of each main time-ambiguity model. For one variant, two separate β-parameters were estimated depending on whether the time-ambiguity range included a possible today-delivery or not, and for the other variant we only estimated a β-parameter when the time-ambiguity ranges did not include a possible today-delivery. In principle, this means we could re-run all our previous time-ambiguity models, but to limit ourselves we decided to only work with the best fitting main six time-ambiguity models, i.e., the best fitting variant of the time perception model, exponential model, additive model, multiplicative model, noise model, and the new crossover model. Given the two different versions we wanted to run per model, this resulted in 12 additional models (thus, plus the two cross-over models, 14 new models in total).

To estimate starting values for the *mle2* function and to reduce the risk of local minima in estimated parameters, we always used a grid-search procedure and selected the 25 best-fitting values (as indicated by lowest G2 values) plus 25 randomly selected values. Each of those 50 starting values was then used in the *mle2* function to fit the data. Of those 50 fits, we then selected the parameters that provided the best fit to the observed data, as indicated by the lowest Bayesian Information Criterion (BIC; Schwarz, 1978). Fits across computational models were compared using mean BIC (i.e., averaged across participants' individual BIC fits): the model with the lowest mean BIC indicated the best fit. Furthermore, we calculated the accuracy of the models by calculating the average probability by which each model predicted the observed choices correctly per participant (i.e., based on the model parameters and the actual choice, the probability of correctly predicting the choice per trial is PrLLchoice if the LL was chosen, and 1 - PrLLchoice if the SS was chosen).

To check whether the estimated parameters (and/or models) were sensible, we performed four different, preregistered sanity checks on the combined data of Study 1 and Study 2. First, we excluded participants without enough variation in their choices (preregistered criterion: >5 SS/LL choices; *n* = 3 in both Study 1 and Study 2, thus *n*=6 in total), because this results in multiple possible solutions in parameter space. For the second sanity check, we calculated a difference score—subtracting the proportion of LL choices during time-ambiguous trials from the proportion of LL choices during time-exact trials—as a rough indicator of time-ambiguity preferences (with positive scores reflecting more LL choices during time-exact than time-ambiguous trials, i.e., time-ambiguity aversion; values close to zero reflecting no difference between time-exact and time-ambiguous trials, i.e., time-ambiguity neutrality; and negative scores reflecting more LL choices during time-ambiguous than time-exact trials, i.e. time-ambiguity seeking). We then computed Kendall’s tau rank-order correlations (given the presence of non-normal distributions) between this difference score and the estimated βSV-values in the whole sample (*N* = 227[[1]](#footnote-1)). We checked whether these rank-order correlations were significant and in the expected direction (i.e., depending on the model, higher βSV-values indicates more time-ambiguity aversion or more time-ambiguity seeking) to keep a model in the formal model comparison. We reasoned that, even if some of the values would be estimated wrongly, the absence (or wrong direction) of a correlation between βSV-values and the difference score with a sample of *N*=227 would reflect something amiss at too big a scale to make it worth adjusting and keeping the model. This check resulted in excluding 5 computational models in total, including one of the previous winning models from Ikink et al. (2019), namely the time perception model with two different parameters to account for the time-ambiguity effect, i.e., βSV and βnoise (with *M*accuracy (excluded models)=.794 vs. *M*accuracy (included models)=.889). Third, we checked at the participant level whether the estimated βSV-values seemed to be estimated correctly. To do so, we plotted and regressed the βSV-values against the difference scores using a robust regression method (i.e., lmrob() from the package robustbase; Maechler et al., 2018), and checked which estimates did not fall within 3 SDs from the regression line. We removed such participants (*n* = 17) from all models in order to keep the model comparisons comparable. On average, these excluded participants were outliers in 6.7 of the 21 to-be-checked models (with *M*accuracy (excluded participants)=.750 vs. *M*accuracy (included participants)=.894). Both accuracy checks confirmed that the excluded models and participants were fitted somewhat worse compared to the included models and participants.

Please note that both the second and third sanity check (i.e., the rank-order correlations and outlier checks – with the latter being done after the former) could only be done for models that have an estimated time-ambiguity parameter which impacts subjective value computations, i.e., not for the standard discounting models (*n*=2) and noise models (where time-ambiguity does not influence subjective value but only the noise parameter θ in the logistic choice rule; *n*=6). Furthermore, they could also not be performed on the crossover models, as time-ambiguity does not follow a linear relationship in these models and higher βSV-values do not indicate more time-ambiguity aversion (*n*=4). Thus, from the 38 models, we checked for 26 models whether they showed a significant rank-order correlation in the expected direction (resulting in the exclusion of 5 models), and for participants from the 21 remaining models we checked whether their estimated βSV-value was an outlier (as this suggests an incorrectly estimated time-ambiguity preference).

The last and fourth preregistered sanity check that we performed on *all* computational models was whether we could predict the choices of participants significantly above chance level (i.e., > 58.1% of choices correctly predicted). For one participant this was not the case in all tested models (with *M*accuracy =.555), thus this participant was also removed. Lastly, it turned out that for 15 participants, three of the tested models could not return a model fit (i.e., negative numbers cannot be raised to a fractional power, which can be a problem for models that include parameter *s* – this happened when applying some of the estimated *k*- and βSV-values; for 5 participants this happened in 2 of the 3 models, for 10 participants in 1 of the 3 models only). To keep the model comparisons comparable, we also removed these 15 participants from all models. Thus, in total we removed 33 participants (15%), of which 9 were from Study 1, and 24 from Study 2. We preregistered that if we excluded participants in the modeling, we would also exclude them in the statistical choice models, given that non-fitting participants might have completed the choice task somehow differently than participants whom we could model. Note that our exclusion criteria were rather strict, because the data of online experiments can be of somewhat lower quality compared to that of lab-based experiments (see e.g., Gould, Cox, Brumby, & Wiseman, 2015). However, since this resulted in a rather high number of excluded participants, we repeated all our analyses with and without those excluded participants and found that none of our main conclusions changed.

With model fits of the remaining 194 participants, we computed the mean BIC per model and compared them. The results showed that from all models, the probabilistic hyperbolic discounting model (Eq. 2), which does not incorporate time-ambiguity, provided overall the best fit in terms of lowest BIC (see Table A; v2). However, fits were overall very similar across many models, with high accuracy (around 90% accurately predicted choices). This seems to indicate that the models made highly similar predictions for the current data and stimuli. We checked whether the fits were different depending on display version (word/timeline), but this was not the case (overall, the model fits were slightly better for the timeline compared to the word version). Furthermore, also when using the whole sample (of *N*=227) or when comparing model fits separately for Study 1 and 2, the results did not substantially change. Thus, these results seem robust.

We also computed the mean rank-order of the models based on individual-participant fits, i.e., indicating per participant which model fitted his/her data best (based on BIC). This indicated that the two standard discounting models had the lowest average rank (probabilistic hyperbolic discounting model: *M*rank=8.433; standard hyperbolic discounting model: *M*rank=9.139), whereas all time-ambiguity models had a mean rank above 11. When looking only at time-ambiguity models, it seemed that one of the exponential models (Table A; v1f) had the lowest mean rank, although a time perception model (Table A; v2a) was the best-fitting among the time-ambiguity models in terms of mean BIC. This again seems to indicate that the model fits were very similar, making it difficult to establish a clear winning time-ambiguity model. Yet even when directly comparing the probabilistic hyperbolic discounting model with the best-fitting time-ambiguity model (here defined by lowest mean BIC, i.e., the time perception model; Table A; v2a), the probabilistic hyperbolic discounting model clearly fitted more participants better compared to the time perception model (probabilistic hyperbolic discounting model: *M*rank=1.237, time perception model: *M*rank=1.763).

Thus, none of the time-ambiguity models improved model fit compared to a standard discounting model that did not account for time-ambiguity. All time-ambiguity models have at least one parameter more, making them per definition more complex. Similar BICs therefore indicate that the increased complexity does not hurt, but neither does it add much (AIC, which is known to penalize less for model complexity, accordingly indicated some time-ambiguity models as slightly better fitting). Instead, all models – regardless of how they incorporated time-ambiguity – fitted the data quite similarly despite different underlying mechanisms. Future studies could therefore try to use stimuli that result in more variation in subjective value, or create specific stimuli for which subjective value predictions clearly differ between the to-be-compared models.

Furthermore, it might be interesting to explore other choice models than alternative-based hyperbolic discounting models, such as, e.g., prospect theory models, attribute-based models, or (attentional) drift diffusion models (although the latter would require that response times and/or measures of visual attention are collected). For example, a time-ambiguity range might be treated as a non-uniform probability distribution, such that depending on an individual’s time-ambiguity preference the probability distribution becomes skewed. Or perhaps some form of weighing or trading-off of not only money and delay differences, but also time-ambiguity differences between choice options occurs. Alternatively, time-ambiguity might result in a bias in the choice process, and/or change the drift rate and delay in the onset of evidence accumulation.

Lastly, although our goal was to find a *general* model that could explain the time-ambiguity effects at the aggregate level, it might also be sensible to take into account individual differences, for example by looking per participant which of the models predicted his/her data best. After all, although *across* participants the crossover and possible-today delivery effects were found, participants might not necessarily display these effects at the single-participant level – which is the level at which the computational models were fitted. Thus, if participants show different time-ambiguity effects, different computational models should best capture that behavior. And indeed, when selecting the lowest BIC-value per participant across all 38 computational models, model fit was somewhat better (*M*best-BIC-per-participant (across all models)=63.811) compared to picking one ‘fixed’ winning model over all participants, i.e., the probabilistic hyperbolic discounting model (*M*BIC-winning-model-overall=69.659). Furthermore, for 75 out of the 194 participants (38.7%) the two standard discounting models provided the best fit, whereas the remaining 119 participants (e.g., 61.3%) were better fitted by a model incorporating time-ambiguity (see also Table A). When selecting the best model per participant across all time-ambiguity models only, we found again a somewhat better model fit compared to both the best-fitting ‘fixed’ standard discounting model and when selecting the best model per participant across the two standard discounting models (*M*best-BIC-per-participant (across time-ambiguity models only)=64.649 vs. *M*BIC-winning-model-overall=69.659 and *M*best-BIC-per-participant (across standard discounting models only)=67.559). Thus, when allowing for individual differences, the time-ambiguity models did actually result in a better fit, although BIC-differences were rather small. Note that this was not preregistered, and as such explorative. Also, some of these differences might be due to fitting noise or overfitting (as we compared fits across 36 versus 2 models). Nevertheless, this seems to indicate that although our main goal was to find a general model across all participants, it might be more suitable to account for individual differences in how participants responded to time-ambiguity.

In conclusion, we explored 36 time-ambiguity models, all incorporating time-ambiguity effects in a somewhat different manner, but the modeling results showed that across all participants, the standard hyperbolic discounting model (not incorporating time-ambiguity in any way) provided the best fit. As we found clear and consistent time-ambiguity effects in our choice models, this seems to indicate that within the framework of discounting models, we could not identify the correct time-ambiguity model to account for these effects. However, when allowing for individual differences in which model fitted each participant best, the different time-ambiguity models together provided a better fit compared to the standard hyperbolic discounting models (both in terms of BIC and number of best-fitted participants). This suggest individual variability in how participants responded to time-ambiguity not only in terms of liking or disliking it, but also with regard to the mechanism via which time-ambiguity might become liked or disliked.

Table A

*Overview of Mean Model Fits for the Tested Computational Models across the Whole Sample (N=194)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | *M*AIC | *M*BIC | *M*Accuracy | *Nr of pps best fitted* |
| *Standard hyperbolic discounting models (not incorporating time-ambiguity):* |  |  |  |  |
|  - v1 (standard hyperbolic discounting model): SV = OV / (1 + k \* D)  | 64.011 | 70.104 | .884 | 50 |
|  **- v2 (probabilistic hyperbolic discounting model): SV = OV / (1 + k \* D)s**  | **60.519** | **69.659** | **.897** | **25** |
| *Time perception models:*  |  |  |  |  |
|  - v1a: incorporating A: SV = OV / (1 + *k* \* (D + βSV \* A)) | 62.305 | 71.445 | .891 | 9 |
|  - v2a: incorporating A: SV = OV / (1 + *k* \* (D + βSV \* A))s  | 58.595 | 70.781 | .904 | 5 |
|  - v1b/v2b: both models using presence/absence A were removed  | *-* | *-* | *-* | *-* |
| *Additive models:*  |  |  |  |  |
|  - v1c: incorporating A: SV = OV / (1 + k \* D) - βSV \* A | 62.752 | 71.892 | .890 | 5 |
|  - v2c: incorporating A: SV = OV / (1 + k \* D)s - βSV \* A | 59.104 | 71.290 | .903 | 4 |
|  - v1d: using presence/absence A (if A, SV = OV / (1 + k \* D) - βSV; else v1) | 62.409 | 71.549 | .892 | 4 |
|  - v2d: using presence/absence A (if A, SV = OV / (1 + k \* D)s - βSV; else v2) | 60.921 | 73.108 | .899 | 0 |
| *Exponential models:* |  |  |  |  |
|  - v1e: incorporating A: SV = OV / (1 + *k* \* D(1 + βsv \* A)) | 62.579 | 71.719 | .891 | 2 |
|  - v2e: incorporating A: SV = OV / (1 + *k* \* D(1 + βsv \* A))s | 58.905 | 71.092 | .903 | 1 |
|  - v1f: using presence/absence A (if A, SV = OV / (1 + *k* \* D(1 + βsv)); else v1) | 61.977 | 71.117 | .894 | 5 |
|  - v2f: using presence/absence A (if A; SV = OV / (1 + *k* \* D(1 + βsv))s; else v2) | 61.257 | 73.444 | .899 | 2 |

*Note*: A = time-ambiguity level; *k* = discount rate; θ = choice stochasticity/noise parameter; βSV = time-ambiguity preference; βnoise = noise-influencing parameter. Models were removed if not passing the model-checks. The best-fitting model (based on lowest BIC) is printed in bold.

Table A *(continued)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | *M*AIC | *M*BIC | *M*Accuracy | *Nr of pps best fitted* |
| *Multiplicative models:*  |  |  |  |  |
|  - v1g: incorporating A: SV = OV / (1 + *k* \* D) \* (1 + βSV \* A)  | 62.444 | 71.584 | .891 | 9 |
|  - v2g: incorporating A: SV = OV / (1 + *k* \* D)s \* (1 + βSV \* A) | 58.781 | 70.968 | .904 | 3 |
|  - v1h: using presence/absence A (if A, SV = OV / (1 + *k* \* D) \* (1 + βSV); else v1)  | 62.356 | 71.496 | .892 | 6 |
|  - v2h: using presence/absence A (if A: SV = OV / (1 + *k* \* D)s \* (1 + βSV); else v2) – removed  | *-* | *-* | *-* | *-* |
| *Noise models:* |  |  |  |  |
| - v1i: incorporating A; only logistic choice rule adjusted (not value function)* formula v1 + Pr(LLchoice) = 1 / (1 + exp(-θ \* (1 + βnoise \* A) \* (SVSS – SVLL)))
 | 65.298 | 74.438 | .885 | 0 |
| - v2i: incorporating A; only logistic choice rule adjusted (not value function)* formula v2 + Pr(LLchoice) = 1 / (1 + exp(-θ \* (1 + βnoise \* A) \* (SVSS – SVLL)))
 | 61.972 | 74.159 | .897 | 0 |
| - v1j: using presence/absence A; only logistic choice rule adjusted (not value function)* if A; formula v1 + Pr(LLchoice) = 1 / (1 + exp(-θ \* (1 + βnoise) \* (SVSS – SVLL)))
* else: formula v1 + Pr(LLchoice) = 1 / (1 + exp(-θ \* (SVSS – SVLL)))
 | 64.421 | 73.561 | .888 | 9 |
| - v2j: using presence/absence A; only logistic choice rule adjusted (not value function)* if A: formula v2 + Pr(LLchoice) = 1 / (1 + exp(-θ \* (1 + βnoise) \* (SVSS – SVLL)))
* else: formula v2 + Pr(LLchoice) = 1 / (1 + exp(-θ \* (SVSS – SVLL)))
 | 61.391 | 73.578 | .899 | 3 |
| *Winning models Ikink et al. (2019):* |  |  |  |  |
| - Time perception model including noise parameter – removed (combining v1a & v1j) | *-* | *-* | *-* | *-* |
| - Additive model including noise parameter:* SV = OV / (1 + k \* D) – βSV \* A (same as v1c)
* if A; Pr(LLchoice)= 1 / (1 + exp(-θ \* (1 + βnoise) \* (SVSS – SVLL))) (same as v1j; using if/else)
 | 62.702 | 74.889 | .895 | 1 |

*Note*: A = time-ambiguity level; *k* = discount rate; θ = choice stochasticity/noise parameter; βSV = time-ambiguity preference; βnoise = noise-influencing parameter. Models were removed if not passing the model-checks. The best-fitting model (based on lowest BIC) is printed in bold.

Table A *(continued)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Model | *M*AIC | *M*BIC | *M*Accuracy | *Nr of pps best fitted* |
| *Crossover models:* |  |  |  |  |
| - v3k: incorporating A: SV = OV \* exp(-1 \* (k\*D)(1 + βsv \* A/D) | 67.453 | 76.593 | .883 | 16 |
| - v3l: using presence/absence A: SV = OV \* exp(-1 \* (k\*D)(1 + βsv) | 64.093 | 73.233 | .891 | 6 |
| *Possible today-delivery models:* |  |  |  |  |
| - Time perception model v2a; estimating onlyβnoPTD | 61.734 | 73.920 | .900 | 2 |
| - Time perception model v2a; estimating βnoPTD and βPTD | 59.230 | 74.464 | .907 | 1 |
| - Additive model v2c; estimating onlyβnoPTD | 61.448 | 73.634 | .899 | 2 |
| - Additive model v2c; estimating βnoPTD and βPTD | 59.581 | 74.814 | .906 | 0 |
| - Exponential model v1f; estimating onlyβnoPTD – removed  | *-* | *-* | *-* | *-* |
| - Exponential model v1f: estimating βnoPTD and βPTD | 61.104 | 73.290 | .898 | 1 |
| - Multiplicative model v2g; estimating onlyβnoPTD | 61.134 | 73.320 | .900 | 0 |
| - Multiplicative model v2g; estimating βnoPTD and βPTD | 58.973 | 74.207 | .908 | 2 |
| - Noise model v1j; estimating onlyβnoise\_noPTD | 63.575 | 72.715 | .889 | 5 |
| - Noise model v1j; estimating βnoise\_noPTD and βnoise\_PTD | 61.081 | 73.268 | .899 | 0 |
| - Crossover model v3l; estimating one βfor trials with no PTD (βnoPTD) | 65.164 | 74.304 | .889 | 15 |
| - Crossover model v3l; estimating two separate β’s for trials with and without PTD (βnoPTD; βPTD) | 63.781 | 75.968 | .895 | 1 |

*Note*: A = time-ambiguity level; *k* = discount rate; θ = choice stochasticity/noise parameter; βSV = time-ambiguity preference; βnoise = noise-influencing parameter. Models were removed if not passing the model-checks. The best-fitting model (based on lowest BIC) is printed in bold.

Table B

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | WAIC Study 1 |  | WAIC Study 2 |  | WAIC Combined |
| Time-ambiguity encoding | ARIM *(n=76)* | Incl. odd trial (*n*=76) | All pps (*n*=88) |  | ARIM *(n=118)* | Incl. odd trial (*n*=118) | All pps (*n*=145) |  | ARIM *(n=194)* | Incl. odd trial (*n*=194) | All pps (*n*=233) |
| Presence/absence of time-ambiguity | 5017 | 5247 | 5888 |  | 7573 | 7959 | 9871 |  | 12564 | 13139 | 15720 |
| Relative time-ambiguity (in %) | 4897 | 5168 | 5705 |  | 7315 | 7818 | 9507 |  | 12198 | 12981 | 15206 |
| Absolute time-ambiguity (in days) | 4858 | 5142 | 5646 |  | 7166 | 7712 | 9364 |  | 11995 | 12831 | 14967 |

 *Comparison of Statistical Model Fit (WAIC) using the As-Reported-In-Manuscript (ARIM) sample versus when either including the One Odd Timeline trial, or All Participants (pps)*

*Note*. Lower WAIC indicates better model fit (but only directly comparable given similar input data). Thus, we show that absolute ambiguity provides the best model fit (across Study 1, Study 2, and the combined sample), and that this was true for the sample as we reported in the manuscript, but also when the odd timeline trial was included (see also SM Fig. A), and when the sample included all participants.

Table C

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Results Study 1a |  | Results Study 2a |  | Results Combineda |  |
| Effects | ARIM *(n=76)* | Incl. odd trial (*n*=76) | All pps(*n*=88) |  | ARIM *(n=118)* | Incl. odd trial (*n*=118) | All pps (*n*=145) |  | ARIM *(n=194)* | Incl. odd trial (*n*=194) | All pps (n=233) |
| Display (word/timeline) | s | s | ns\* |  | s | s | s |  | s | s | s |
| Amount  | s | s | s |  | s | s | s |  | s | s | s |
| Delay (midpoint) | s | s | s |  | s | s | s |  | s | s | s |
| Time-ambiguity (absolute; in days) | ns | t\* | ns |  | ns | ns | ns |  | ns | ns | ns |
| Display x Amount  | t | ns\* | s\* |  | s | ns\* | t\* |  | s | ns\* | s |
| Display x Delay  | ns | t\* | ns |  | t | s\* | t |  | s | s | t\* |
| Display x Time-ambiguity | ns | ns | ns |  | ns | ns | ns |  | ns | ns | ns |
| Time-ambiguity x Amount | s | s | s |  | ns | ns | ns |  | s | ns\* | s |
| Time-ambiguity x Delay  | s | s | s |  | s | s | s |  | s | s | s |
| Amount x Delay  | ns | ns | ns |  | ns | s\* | ns |  | ns | s\* | ns |
| Time-ambiguity x Amount x Display | ns | ns | ns |  | ns | s\* | ns |  | ns | t\* | ns |
| Time-ambiguity x Delay x Display | s | s | t\* |  | ns | t\* | ns |  | s | s | s |
| Amount x Delay x Display | t | ns\* | t |  | ns | ns | ns |  | t | ns\* | t |
| Time-ambiguity x Amount x Delay  | ns | ns | ns |  | s | s | s |  | s | t\* | s |
| Time-ambiguity x Amount x Delay x Display | ns | ns | ns |  | s | t\* | ns\* |  | s | s | t\* |

 *Comparison of Results of the Main Choice Model using the As-Reported-In-Manuscript (ARIM) sample versus when either including the One Odd Timeline Trial, or All Participants (pps) (with \* indicating a change relative to results of the ARIM sample)*

*Note*: a pps = participants; s = significant; t = trend; ns = non-significant. Thus, some of the results changed when including the odd timeline trial or when including all participants (compared to the ARIM sample). It seems overall more changes in results occurred after including the odd timeline trial (i.e., specifically interactions involving amount) compared to when including all participants, consistent with the fact that the choices in that specific trial do not fit the rest of the choice patterns (see also SM Fig. A). Thus, we feel that the exclusion of this trial is appropriate. Importantly, including all participants did not change any of our main conclusions regarding either the display or time-ambiguity effects.

Table D

*Comparison of Results of the Possible Today Delivery (PTD) model using the As-Reported-In-Manuscript (ARIM) sample versus All Participants (with \* indicating a change relative to results of the ARIM sample)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Effects | Results Study 1a |  | Results Study 2a |  | Results Combineda  |
| ARIM *(n=76)* | All pps (*n*=88) |  | ARIM *(n=118)* | All pps (*n*=145) |  | ARIM *(n=194)* | All pps (n=233) |
| Display (word/timeline) | s | s |  | s | s |  | s | s |
| Amount  | s | s |  | s | s |  | s | s |
| Delay (midpoint) | s | s |  | s | s |  | s | s |
| Time-ambiguity (absolute; in days) | s | s |  | t | s\* |  | s | s |
| Display x Amount  | s | s |  | s | s |  | s | s |
| Display x Delay  | ns | ns |  | s | s |  | s | s |
| Display x Time-ambiguity | ns | ns |  | ns | ns |  | ns | ns |
| Time-ambiguity x Amount | s | s |  | ns | ns |  | ns | ns |
| Time-ambiguity x Delay  | s | s |  | s | s |  | s | s |
| Amount x Delay  | ns | ns |  | s | s |  | s | s |
| Time-ambiguity x Amount x Display | ns | ns |  | ns | ns |  | ns | ns |
| Time-ambiguity x Delay x Display | s | s |  | t | t |  | s | s |
| Amount x Delay x Display | ns | ns |  | s | s |  | s | s |
| Time-ambiguity x Amount x Delay  | ns | t\* |  | ns | ns |  | ns | ns |
| Time-ambiguity x Amount x Delay x Display | ns | t\* |  | ns | ns |  | t | s\* |
| Possible today-delivery (yes/no) | s | s |  | s | s |  | s | s |
| Possible today-delivery x Time-ambiguity | t | s\* |  | t | s\* |  | s | s |

*Note*: a pps = participants; s = significant; t = trend; ns = non-significant. No results for also including the one odd timeline trial are reported, as this restricted dataset never included that trial. Thus, some of the results changed when including all participants (compared to the ARIM sample). Most importantly, though, none of these changes in results change our main conclusions regarding the possible today-delivery (PTD) effect. If anything, the effects of interest (time-ambiguity, PTD, and their interaction) seem to be somewhat stronger in the full sample.

Table E

*Overview of the 39 Unique Trials based on Delay midpoint and Time-Ambiguity Level*

|  |  |  |
| --- | --- | --- |
| Trial | Delay midpoint | Time-ambiguity coding |
| Presence/absence | Absolute (in days) | Relative (in %) |
| 1 day | 1 | absent | 0 | 0 |
| 0 to 2 days | 1 | present | 2 | 200 |
| 10 days | 10 | absent | 0 | 0 |
| 9 to 11 days | 10 | present | 2 | 20 |
| 8 to 12 days | 10 | present | 4 | 40 |
| 5 to 15 days | 10 | present | 10 | 100 |
| 0 to 20 days | 10 | present | 20 | 200 |
| 30 days | 30 | absent | 0 | 0 |
| 29 to 31 days | 30 | present | 2 | 6.666 |
| 28 to 32 days | 30 | present | 4 | 13.333 |
| 27 to 33 days | 30 | present | 6 | 20 |
| 25 to 35 days | 30 | present | 10 | 33.333 |
| 24 to 36 days | 30 | present | 12 | 40 |
| 20 to 40 days | 30 | present | 20 | 66.667 |
| 15 to 45 days | 30 | present | 30 | 100 |
| 0 to 60 days | 30 | present | 60 | 200 |
| 90 days | 90 | absent | 0 | 0 |
| 89 to 91 days | 90 | present | 2 | 2.222 |
| 88 to 92 days | 90 | present | 4 | 4.444 |
| 85 to 95 days | 90 | present | 10 | 11.111 |
| 81 to 99 days | 90 | present | 18 | 20 |
| 80 to 100 days | 90 | present | 20 | 22.222 |
| 72 to 108 days | 90 | present | 36 | 40 |
| 60 to 120 days | 90 | present | 60 | 66.667 |
| 45 to 135 days | 90 | present | 90 | 100 |
| 0 to 180 days | 90 | present | 180 | 200 |
| 180 days | 180 | absent | 0 | 0 |
| 179 to 181 days | 180 | present | 2 | 1.111 |
| 178 to 182 days | 180 | present | 4 | 2.222 |
| 175 to 185 days | 180 | present | 10 | 5.5556 |
| 170 to 190 days | 180 | present | 20 | 11.111 |
| 162 to 198 days | 180 | present | 36 | 20 |
| 150 to 210 days | 180 | present | 60 | 33.333 |
| 144 to 216 days | 180 | present | 72 | 40 |
| 120 to 240 days | 180 | present | 120 | 66.667 |
| 90 to 270 days | 180 | present | 180 | 100 |
| 0 to 360 days | 180 | present | 360 | 200 |
| 270 days | 270 | absent | 0 | 0 |
| 360 days | 360 | absent | 0 | 0 |



*Figure A*. Overview of proportion LL choices in the timeline and word version per study (aggregated across participants), for LL amount €5.20 and exact delays between 1 and 360 days. The ‘odd’ removed trial in the timeline version is visible (encircled in black) in both Study 1 and Study 2, i.e., €5.20 in 180 days. Since only the data in the timeline version showed this unexpected increase in LL choice for a delay of 180 days (compared to both shorter and longer delays), we did not remove that trial in the word version. Unfortunately, our license to Unipark Questback expired before we could check what exactly went wrong. Our best guess is a programming error, such that e.g. a different LL amount was shown.

Table F

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the Main effect of Time-Ambiguity per Short and Long Delay*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time-ambiguity effect |  | Overview effect per modela |
|  | Study 1 (*n*=76) |  | Study 2 (*n*=118) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  | B | 95% CI |  |
| Short delays (1/10/30 days) | -0.713 | [-1.361, -0.070] |  | -0.280 | [-0.917, 0.428] |  | -0.565 | [-1.090, -0.051] |  | t / ns / s |
| Long delays (90/180 days) | 0.025 | [-0.219, 0.264] |  | 0.189 | [0.026, 0.352] |  | 0.155 | [0.011, 0.302] |  | ns / t / s |

a s = significant; t = trend effect; ns = non-significant. For trend effects, 90% posterior CIs are reported.

Table G

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the Main effect of Time-Ambiguity per LL Amount*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time-ambiguity effect |  | Overview effect per modela |
|  | Study 1 (*n*=76) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  |
| Amount €5.20 | -0.580 | [-5.295, 2.895] |  | 1.297 | [-1.795, 3.965] |  | ns / ns |
| Amount €10.50 | -0.108 | [-0.663, 0.416] |  | 0.156 | [-0.128, 0.433] |  | ns / ns |
| Amount €16.80 | -0.341 | [-0.823, 0.131] |  | 0.009 | [-0.291, 0.334] |  | ns / ns |
| Amount €25.30  | -0.257 | [-0.659, 0.179] |  | -0.290 | [-0.633, 0.093] |  | ns / ns |

a ns = non-significant

Table H

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the Time-Ambiguity by Delay interaction (i.e., the Crossover Effect) per LL Amount*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time-ambiguity by Delay (i.e., crossover) effect |  | Overview effect per modela |
|  | Study 2 (*n*=118) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  |
| Amount €5.20 | 2.274 | [-0.983, 5.633] |  | 3.540 | [0.224, 6.038] |  | ns / s |
| Amount €10.50 | 0.494 | [0.151, 0.823] |  | 0.538 | [0.225, 0.845] |  | s / s |
| Amount €16.80 | 0.193 | [-0.334, 0.619] |  | 0.448 | [0.108, 0.754] |  | ns / s |
| Amount €25.30  | 0.161 | [-0.458, 0.663] |  | 0.240 | [-0.157, 0.577] |  | ns / ns |

a s = significant; ns = non-significant

Table I

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the LL Amount effect per Display Version*

|  |  |  |  |
| --- | --- | --- | --- |
|  | LL amount effect |  | Overview effect per modela |
|  | Study 1 (*n*=76) |  | Study 2 (*n*=118) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  | B | 95% CI |  |
| Timeline version | 4.904 | [4.105, 5.744] |  | 6.406 | [5.214, 7.723] |  | 5.687 | [4.924, 6.463] |  | s / s / s |
| Word version | 3.841 | [2.516, 5.225] |  | 4.355 | [3.427, 5.364]] |  | 4.096 | [3.264, 4.909] |  | s / s / s |

a s = significant

Table J

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the LL Delay (Midpoint) Effect per Display Version*

a s = significant.

|  |  |  |  |
| --- | --- | --- | --- |
|  | LL Delay effect |  | Overview effect per modela |
|  | Study 2 (*n*=118) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  |
| Timeline version | -3.134 | [-4.309, -1.979] |  | -3.477 | [-4.349, -2.645] |  | s / s  |
| Word version | -4.384 | [-5.290, -3.483] |  | -4.523 | [-5.297, -3.784] |  | s / s |

**Appendix C: Details of the interaction effects between display and time-ambiguity**

We expected that display and time-ambiguity might show a 2-way interaction, such that the time-ambiguity effect would be more pronounced in the timeline than the word version (because the timeline version shows a visually salient box to indicate the time-ambiguity range). However, our results clearly indicated this was not the case (non-significant 2-way interaction in all 3 samples; see Table 2). Instead, we found a 3-way interaction in Study 1 and the combined sample (similar direction but non-significant in Study 2), showing that the time-ambiguity by delay interaction (i.e., the crossover effect, with time-ambiguity aversion across shorter delays and time-ambiguity liking across longer delays) was moderated by display. Follow-up analyses per version showed that the crossover effect was significant for both display versions, but stronger in the word version (SM Table K).

Furthermore, we found an unexpected 4-way interaction in Study 2 and the combined sample, which indicated that the 3-way interaction between time-ambiguity, delay midpoint, and amount magnitude was larger in the timeline than the word version: While the 3-way interaction was significant across the two display versions (this interaction suggests that the crossover effect becomes weaker for larger LL amounts; see SM Table H), it was non-significant in either version separately. However, based on the estimated coefficients for the 3-way interaction per version (see also SM Table L), it seems that the significance of the 3-way interaction across both versions was mostly driven by the timeline version.

Thus, although the crossover effect was stronger in the word version, the crossover effect also became smaller with larger amounts, and this 3-way interaction was stronger in the timeline version than the word version. Thus, while two of the time-ambiguity (interaction) effects differed per display version, our hypothesis that specifically the timeline version would show more pronounced time-ambiguity effects than the word version (H5) was not confirmed.

Table K

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the Time-Ambiguity by Delay interaction (i.e., the Crossover Effect) per Display Version*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time-ambiguity by Delay (i.e., cross-over) effect |  | Overview effect per modela |
|  | Study 1 (*n*=76) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  |
| Timeline version | 0.377 | [0.043, 0.706] |  | 0.522 | [0.233, 0.796] |  | s / s |
| Word version | 0.958 | [0.481, 1.455] |  | 0.715 | [0.400, 1.022] |  | s / s |

a s = significant.

Table L

*Estimated Coefficients (Bs) and 95% Credible Intervals (CIs) for the 3-way interaction between Time-Ambiguity, Delay Midpoint, and LL amount per Display Version*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time-ambiguity by Delay by LL amount effect |  | Overview effect per modela |
|  | Study 2 (*n*=118) |  | Combined (*n*=194) |  |
| Follow-up model | B | 95% CI |  | B | 95% CI |  |
| Timeline version | -0.258 | [-0.753, 0.157] |  | -0.123 | [-0.439, 0.161] |  | ns / ns |
| Word version | -0.056 | [-0.458, 0.289] |  | -0.069 | [-0.389, 0.223] |  | ns / ns |

a ns = non-significant.

Table M

*Proportion of LL choices per delay midpoint, time-ambiguity level, amount, and display version in the combined sample (N=194)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Delay | Ambig. level | p (LL choices) in Word Version |  | p (LL choices) in Timeline version |
| amount €5.20 | amount €10.50 | amount €16.80 | amount €25.30 |  | amount €5.20 | amount €10.50 | amount €16.80 | amount €25.30 |
| 1 day | 0 days | 0.616 | 1.00 | 1.00 | 0.989 |  | 0.800 | 1.00 | 1.00 | 1.00 |
|  | 2 days | .0.475 | 1.00 | 1.00 | 1.00 |  | 0.695 | 1.00 | 1.00 | 0.989 |
|  |  |  |  |  |  |  |  |  |  |  |
| 10 days | 0 days | 0.131 | 0.960 | 0.970 | 0.990 |  | 0.347 | 0.979 | 0.989 | 0.989 |
| 2 days  | 0.141 | 0.990 | 0.970 | 0.980 |  | 0.221 | 0.989 | 0.989 | 1.000 |
| 4 days | 0.121 | 0.949 | 0.980 | 0.980 |  | 0.316 | 0.979 | 0.989 | 1.000 |
|  | 6 days  |  |  |  |  |  |  |  |  |  |
|  | 10 days | 0.131 | 0.929 | 0.980 | 0.990 |  | 0.221 | 0.968 | 1.000 | 0.989 |
|  | 12 days |  |  |  |  |  |  |  |  |  |
|  | 18 days |  |  |  |  |  |  |  |  |  |
|  | 20 days | 0.111 | 0.919 | 0.960 | 0.980 |  | 0.284 | 0.989 | 0.979 | 1.000 |
|  |  |  |  |  |  |  |  |  |  |  |
| 30 days | 0 days | 0.020 | 0.737 | 0.869 | 0.929 |  | 0.095 | 0.853 | 0.947 | 0.958 |
| 2 days | 0.030 | 0.768 | 0.879 | 0.970 |  | 0.137 | 0.853 | 0.926 | 0.989 |
|  | 4 days | 0.020 | 0.717 | 0.869 | 0.929 |  | 0.105 | 0.853 | 0.926 | 0.958 |
|  | 6 days | 0.030 | 0.717 | 0.879 | 0.980 |  | 0.137 | 0.842 | 0.895 | 0.979 |
|  | 10 days | 0.051 | 0.727 | 0.838 | 0.919 |  | 0.084 | 0.842 | 0.895 | 0.937 |
|  | 12 days | 0.040 | 0.747 | 0.879 | 0.929 |  | 0.126 | 0.832 | 0.916 | 0.947 |
|  | 18 days |  |  |  |  |  |  |  |  |  |
|  | 20 days | 0.051 | 0.788 | 0.828 | 0.949 |  | 0.095 | 0.832 | 0.905 | 0.989 |
|  | 30 days | 0.030 | 0.758 | 0.848 | 0.949 |  | 0.116 | 0.842 | 0.916 | 0.968 |
|  | 36 days |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 90 days | 0 days | 0.000 | 0.354 | 0.485 | 0.636 |  | 0.063 | 0.537 | 0.726 | 0.863 |
| 2 days | 0.010 | 0.323 | 0.404 | 0.566 |  | 0.042 | 0.589 | 0.653 | 0.884 |
|  | 4 days | 0.000 | 0.343 | 0.455 | 0.596 |  | 0.063 | 0.474 | 0.663 | 0.842 |
|  | 6 days |  |  |  |  |  |  |  |  |  |
|  | 10 days | 0.000 | 0.374 | 0.475 | 0.576 |  | 0.032 | 0.516 | 0.737 | 0.842 |
|  | 12 days |  |  |  |  |  |  |  |  |  |
|  | 18 days | 0.010 | 0.323 | 0.465 | 0.576 |  | 0.063 | 0.526 | 0.695 | 0.821 |
|  | 20 days | 0.010 | 0.303 | 0.444 | 0.596 |  | 0.032 | 0.579 | 0.611 | 0.821 |
|  | 30 days  |  |  |  |  |  |  |  |  |  |
|  | 36 days | 0.000 | 0.293 | 0.434 | 0.556 |  | 0.063 | 0.495 | 0.695 | 0.853 |
|  | 60 days | 0.000 | 0.354 | 0.475 | 0.697 |  | 0.032 | 0.547 | 0.663 | 0.842 |
|  | 72 days  |  |  |  |  |  |  |  |  |  |
|  | 90 days | 0.010 | 0.303 | 0.424 | 0.606 |  | 0.063 | 0.474 | 0.716 | 0.853 |
|  | 120 days  |  |  |  |  |  |  |  |  |  |
|  | 180 days | 0.030 | 0.404 | 0.515 | 0.626 |  | 0.084 | 0.611 | 0.674 | 0.821 |

*Continuation Table M*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Delay | Ambig. level | p (LL choices) in Word Version |  | p (LL choices) in Timeline version |
| amount €5.20 | amount €10.50 | amount €16.80 | amount €25.30 |  | amount €5.20 | amount €10.50 | amount €16.80 | amount €25.30 |
| 180 days | 0 days  | 0.000 | 0.212 | 0.354 | 0.475 |  | excluded | 0.320 | 0.568 | 0.747 |
| 2 days | 0.000 | 0.242 | 0.283 | 0.475 |  | 0.032 | 0.400 | 0.516 | 0.705 |
|  | 4 days  | 0.010 | 0.263 | 0.293 | 0.434 |  | 0.011 | 0.421 | 0.537 | 0.716 |
|  | 6 days |  |  |  |  |  |  |  |  |  |
|  | 10 days | 0.000 | 0.212 | 0.343 | 0.475 |  | 0.021 | 0.326 | 0.558 | 0.758 |
|  | 12 days |  |  |  |  |  |  |  |  |  |
|  | 18 days |  |  |  |  |  |  |  |  |  |
|  | 20 days | 0.000 | 0.242 | 0.303 | 0.465 |  | 0.011 | 0.442 | 0.516 | 0.684 |
|  | 30 days  |  |  |  |  |  |  |  |  |  |
|  | 36 days | 0.000 | 0.293 | 0.434 | 0.556 |  | 0.063 | 0.495 | 0.695 | 0.853 |
|  | 60 days | 0.000 | 0.212 | 0.303 | 0.495 |  | 0.011 | 0.337 | 0.589 | 0.747 |
|  | 72 days | 0.000 | 0.202 | 0.323 | 0.495 |  | 0.032 | 0.411 | 0.537 | 0.726 |
|  | 90 days |  |  |  |  |  |  |  |  |  |
|  | 120 days | 0.000 | 0.222 | 0.303 | 0.465 |  | 0.021 | 0.389 | 0.495 | 0.684 |
|  | 180 days | 0.000 | 0.263 | 0.313 | 0.505 |  | 0.000 | 0.421 | 0.526 | 0.663 |
|  | 360 days | 0.010 | 0.263 | 0.414 | 0.535 |  | 0.042 | 0.432 | 0.516 | 0.695 |
|  |  |  |  |  |  |  |  |  |  |  |
| 270 days | 0 days | 0.000 | 0.212 | 0.283 | 0.455 |  | 0.021 | 0.337 | 0.453 | 0.632 |
|  |  |  |  |  |  |  |  |  |  |  |
| 360 days | 0 days | 0.000 | 0.202 | 0.273 | 0.374 |  | 0.000 | 0.295 | 0.358 | 0.558 |

**Appendix D: Details of follow-up models for the possible today-delivery model**

As time-ambiguity aversion was found to be significant in the possible today-delivery model (using 56 of the 156 trials; Table 3) but not in the main choice model (using all 156 trials; Table 2; SM Table C), we wanted to check if this was due to using this specific subset of the data. We therefore repeated the analysis once without the possible today-delivery factor and its interaction with time-ambiguity level (thus running the same model as the main choice model in the full dataset – but now in the restricted dataset), and once with possible today-delivery as main effect only. As a third model, we removed all PTD trials (*n*=20) from the full dataset, and reran the main choice model on that dataset.

When running the main choice model in the restricted dataset, time-ambiguity was consistently not significant in any of the samples (Study 1: B = -0.241, 95% CI [-0.591, 0.119]; Study 2: B = 0.135, 95% CI [-0.268, 0.581]; combined sample: B = 0.004, 95% CI [-0.264, 0.287]), indicating that the time-ambiguity aversion we found in the possible today-delivery model was not due to using a subset of the trials. Secondly, when the possible today-delivery was added as main effect only, time-ambiguity was (similar to the original possible today-delivery model we ran) significant in Study 1 and the combined sample, but not in Study 2 (Study 1: B = -0.707, 95% CI [-1.108, -0.301]; Study 2: B = -0.253, 95% CI [-0.672, 0.202]; combined sample: B = -0.391, 95% CI [-0.697, -0.083]). Possible today-delivery as main effect was again significant in all 3 samples (Study 1: B = -0.490, 95% CI [-0.727, -0.248]; Study 2: B = -0.483, 95% CI [-0.692, -0.280]; combined sample: B = -0.465, 95% CI [-0.616, -0.319]). In the third model, where all PTD trials were removed from the full dataset, time-ambiguity was significant in Study 1, and a trend effect in both Study 2 and the combined sample (Study 1: B = -0.232, 95% CI [-0.455, -0.021]; Study 2: B = -0.209, 90% CI [-0.396, -0.019]; combined sample: B = -0.141, 90% CI [-0.278, -0.004]). This shows that possible today-delivery is a reliable effect (i.e., found consistently when added to the model), and that accounting for its effect in the model results in making time-ambiguity a significant (or trend) main effect.

However, in the possible today-delivery model, delay midpoints for PTD-trials are systematically lower than those of the no PTD-trials (see Table 3), making it possible that the PTD effect is partially explained by its confound with delay (note, however, that delay *was* included in the model and as such controlled for). In an attempt to address this confound, we decided to run two different follow-up models to further clarify whether PTD is an independent effect or might be an artifact due to the delay confound. In the first follow-up model (FUM-1), we included trials where the delay midpoints were similar, but time-ambiguity ranges differed (see SM-Table N). Please note that here, time-ambiguity ranges for PTD trials are systematically higher than those of non-PTD trials, i.e., in this set of trials time-ambiguity level is a confound (the price paid for minimizing the delay confound). As so few trials were included (*n*=32), we only tested the main effects of version, delay midpoint, amount, and the PTD factor. In the second follow-up model (FUM-2), we decided to include all time-ambiguous trials (*n*=138), including all PTD-trials (*n*=20) and all no PTD-trials (*n*=108) without trying to match on delay midpoint or time-ambiguity level. We ran the main choice model again, plus an additional possible today-delivery factor (yes/no) as main effect. We pre-registered that if the PTD effect would be significant in both these follow-up models, we would interpret it as a reliable effect, and if in one follow-up model only, a tentative effect.

The results were very consistent across Study 1, Study 2, and the combined sample: in FUM-1 (i.e., similar delay midpoints, but different time-ambiguity levels), the possible today-delivery effect was consistently not significant (Study 1: B = -0.043, 95% CI [-0.212, 0.126];Study 2: B = 0.003, 95% CI [-0.149, 0.160]; combined sample: B = -0.015, 95% CI [-0.123, 0.096]). Yet in FUM-2 (all time-ambiguous trials, not matched on either delay midpoint or time-ambiguity level), possible today-delivery was consistently significant (Study 1: B = -0.432, 95% CI [-0.618, -0.251]; Study 2: B = -0.358, 95% CI [-0.517, -0.202]; combined sample: B = -0.391, 95% CI [-0.509, -0.275]), as was time-ambiguity as main effect (Study 1: B = -0.890, 95% CI [-1.287, -0.503]; Study 2: B = -0.538, 95% CI [-0.918, -0.142]; combined sample: B = -0.641, 95% CI [-0.924, -0.352]).

Given that the PTD effect is impacted by the unavoidable confound between the PTD and delay midpoint predictor, we interpret the PTD effect as a somewhat tentative effect: If delay midpoints are held constant, we do not find the PTD effect (or, alternatively, the difference between the two time-ambiguity levels is not significant). However, we found the PTD effect very consistently in both independent studies (and the combined sample), and adding this factor to the model made time-ambiguity significant as a main effect. Therefore, it seems interesting for future studies to investigate the PTD effect in a more targeted way, for example by using trials where the delay midpoints of PTD and no-PTD trials are better matched.

Table N

*Trials included in Restricted Dataset 2, to Alternatively test the Possible-Today Delivery Hypothesis (follow-up model 1)*

|  |  |
| --- | --- |
| Possible today-delivery  | No possible today-delivery |
| 0 to 20 days | 5 to 15 days  |
| 0 to 60 days | 15 to 45 days  |
| 0 to 180 days  | 45 to 135 days  |
| 0 to 360 days  | 90 to 270 days  |
| *Total:* 4\*4 (amount) = 16 trials | *Total*: 4\*4 (amount) = 16 trials  |

*Note:* here delay midpoints are similar, but absolute time-ambiguity levels differ.

**Appendix E: Overview of the 7 items used in Study 3**

Here we present the 7 items that were used in Study 3. Note that participants first received instructions and check questions (not presented here), to make sure they understood the bag descriptions. Each item tested a specific comparison, and the order in which the items were displayed to participants was randomized. Each item started with the same text:

*“Suppose you have 2 bags, that each contain 100 balls. Each ball contains a number that indicates a specific delivery time.****From one of the two bags, a ball will be drawn at random, which determines when you will receive £100****. For example, if a ball with the number 20 is drawn, you will receive £100 in 20 days from now; if a ball with the number 80 is drawn, you will receive £100 in 80 days from now.”*

After the description of the two bags (see further below), we always asked participants the following: *Please indicate on the scale below to what extent you have a preference for drawing a ball from either bag 1 or bag 2.* The answer scale ranged from -50 (*definitely bag 1*) to 50 (*definitely bag 2*) with the label “*no clear preference*” at a score of 0.

Below we list the actual descriptions that were used for bag 1 and 2. We present here the task version where bag 1 was always the more time-certain; in the other task version this was bag 2 (this was counterbalanced across participants). Note that for participants, the items were not numbered. Also, they did not include a description of the type of comparison, which we added here for ease of understanding. Thus, participants always read the start-text (‘Suppose you have …”), then the descriptions of the 2 bags was given using similar bullet-points as here, and then the rating question (‘Please indicate …”).

* *Item 1*: 1 or 100 days using a 50/50 time-risky versus an unknown time-ambiguous distribution.
* bag 1 contains 50 balls with the number 1 on them, and 50 balls with the number 100 on them.
* bag 2 contains 100 balls that each have either the number 1 or the number 100 on them, but it is unclear how many balls have which number. Thus, number 1 could be in there equally often as number 100, but it could also be in there more often, less often, or not at all.
* *Item 2*: 1 to 100 days using a uniform time-risky versus an unknown time-ambiguous distribution.
	+ bag 1 contains 100 balls that are numbered from 1 to 100. Thus, each number is in there once.
	+ bag 2 contains 100 balls that each have a number between 1 and 100 on them, but it is unclear how many balls have which number. Thus, some numbers could be in there more often than other numbers, and some numbers may not be in there at all.
* *Item 3*: 1 or 100 days using a 50/50 time-risky distribution, versus 1 to 100 days using an unknown time-ambiguous distribution.
	+ bag 1 contains 50 balls with the number 1 on them, and 50 balls with the number 100 on them.
	+ bag 2 contains 100 balls that each have a number between 1 and 100 on them, but it is unclear how many balls have which number. Thus, some numbers could be in there more than other numbers, and some numbers may not be in there at all.
* *Item 4*: a 51 days time-exact option versus 1 or 100 days using a 50/50 time-risky distribution.
	+ bag 1 contains 100 balls that all have the number 51 on them.
	+ bag 2 contains 50 balls with the number 1 on them, and 50 balls with the number 100 on them.
* *Item 5*: a 51 days time-exact option versus 1 or 100 days using an unknown time-ambiguous distribution.
	+ bag 1 contains 100 balls that all have the number 51 on them.
	+ bag 2 contains 100 balls that each have either the number 1 or the number 100 on them, but it is unclear how many balls have which number. Thus, number 1 could be in there equally often as number 100, but it could also be in there more often, less often, or not at all.
* *Item 6*: a 51 days time-exact option versus 1 to 100 days using a uniform time-risky distribution.
	+ bag 1 contains 100 balls that all have the number 51 on them.
	+ bag 2 contains 100 balls that are numbered from 1 to 100. Thus, each number is in there once.
* *Item 7*: a 51 days time-exact option versus 1 to 100 days using an unknown time-ambiguous distribution.
	+ bag 1 contains 100 balls that all have the number 51 on them.
	+ bag 2 contains 100 balls that each have a number between 1 and 100 on them, but it is unclear how many balls have which number. Thus, some numbers could be in there more often than other numbers, and some numbers may not be in there at all.

Table O

*Descriptive statistics of the 7 items used in Study 3 (N=202) – all with range -50 to 50*

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Choice options | *M* (SD) | Median |
| 1 | time-risky (1 or 100d 50/50) vs. time-ambiguous (1 or 100d unknown) | -11.80 (30.225) | -19 |
| 2 | time-risky (1 to 100d uniform) vs. time-ambiguous (1 to 100 unknown) | -11.95 (27.683) | -13.5 |
| 3 | time-risky (1 or 100d 50/50) vs. time-ambiguous (1 to 100d unknown) | -8.837 (29.868) | -10 |
| 4 | time-exact (51d) vs. time-risky (1 or 100d 50/50) | 3.762 (34.516) | 11 |
| 5 | time-exact (51d) vs. time-ambiguous (1 or 100d unknown)  | 1.134 (31.095) | 0 |
| 6 | time-exact (51d) vs. time-risky (1 to 100d uniform)  | 4.663 (33.781) | 12 |
| 7 | time-exact (51d) vs time-ambiguous (1 to 100d unknown) | -2.822 (31.712) | -1.5 |

*Note:* All scores are coded such that negative scores indicate a preference for a time-risky option (for items 1 to 3), or a time-exact option (for items 4 to 7).

References

Ebert, J., & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of

the near and far future. *Management Science*, 53, 1423–1438.

https://doi.org/10.1287/mnsc.1060.0671

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The*

*Journal of Socio-Economics*, *40*, 35-42. https://doi.org/10.1016/j.socec.2010.10.008

Gould, S. J., Cox, A. L., Brumby, D. P., & Wiseman, S. (2015). Home is where the lab is: a

comparison of online and lab data from a time-sensitive study of interruption. *Human Computation*, *2*, 45-67. doi: 10.15346/hc.v2i1.4

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite

effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 418–427.

Ikink, I., Engelmann, J. B., van den Bos, W., Roelofs, K., & Figner, B. (2019). Time

ambiguity during intertemporal decision-making is aversive, impacting choice and neural value coding. *NeuroImage*, *185*, 236-244. https://doi.org/10.1016/j.neuroimage.2018.10.008

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M.,

… & di Palma M. A. (2018). robustbase: Basic Robust Statistics R package version 0.93-3. URL: http://CRAN.R-project.org/package=robustbase

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M.

L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analysis of behavior: Vol. 5. The effect of delay and intervening events on reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.

McKerchar, T. L., & Renda, C. R. (2012). Delay and probability discounting in humans: An

overview. *The Psychological Record*, *62*, 817-834. doi: https://link.springer.com/content/pdf/10.1007/BF03395837.pdf

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting

through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, *66*, 138-148. doi: 10.1016/j.neuron.2010.03.026

Peters, J., Miedl, S. F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal

discounting models in controls and pathological gamblers. *PloS one*, *7*, e47225. https://doi.org/10.1371/journal.pone.0047225

R Core Team (2018). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four score and seven years from

now: The date/delay effect in temporal discounting. *Management Science*, *51*, 1326-1335. https://doi.org/10.1287/mnsc.1050.0412

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic*

*Studies*, *4*, 155-161. https://www.jstor.org/stable/2967612

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-

464. https://projecteuclid.org/euclid.aos/1176344136

Tymula, A., Belmaker, L. A. R., Roy, A. K., Ruderman, L., Manson, K., Glimcher, P.

W., & Levy, I. (2012). Adolescents’ risk-taking behavior is driven by tolerance to

ambiguity. *Proceedings of the National Academy of Sciences, 109,* 17135-17140. doi: 10.1073/pnas.1207144109

1. Note: Initially, the whole sample size included 233 participants. However, participants without enough variation in choice (preregistered criterion: > 5 times the SS/LL option chosen out of 156 trials; *n*=6) were never included in the modeling, as too little variation is known to complicate parameter estimation due to multiple possible solutions. [↑](#footnote-ref-1)