

How Much is Too Much: Validity and Reliability of Time Preferences.

Daniel Wall

Rutgers University

Ye Li

University of California, Riverside

Olivier Toubia

Columbia University

Eric J. Johnson

Columbia University

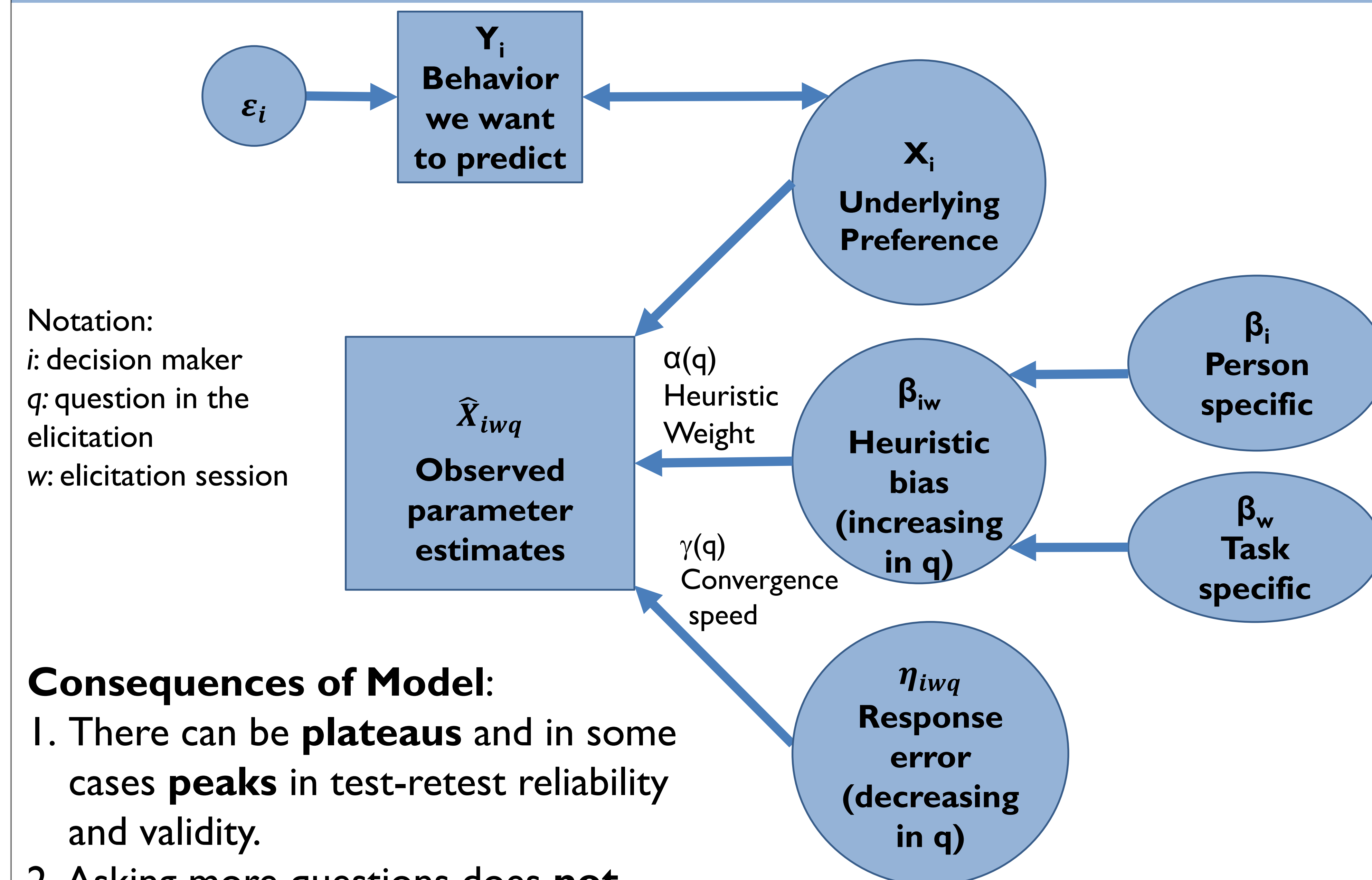
Abstract

When measuring preferences, how many questions should we ask a participant? Psychometrics suggests the more questions the better the estimate. However other research suggests that people develop heuristics, which should decrease reliability and validity, when answering questions. Through a formal model and experimental evidence, we explore how heuristics affect reliability and validity. The model underscores **the need to efficiently and robustly estimate preferences before task and person specific heuristics dominate the participants true preferences.** We test the predictions of the model in two studies. Participants answered a series of intertemporal choices; we find that **test-retest reliability plateaus after 8 of 20 questions and that concurrent and external validity peak after 8 questions.** When measuring preferences, **less can be more.**

Verbal Description of Model

Our model reflects **two countervailing forces** that vary as the number of questions increase. First, **more questions provide more data**, which reduces the standard error of our estimate—i.e., our estimate converges to some value. Second, **more questions may increase the intensity of heuristic use**, which may introduce bias. In other words, as more questions are added, our estimates become more precise (i.e., they have less standard error and converge to some value), but potentially also more biased (i.e., the value to which they converge may be further from the truth).

Conceptual Model



Consequences of Model:

1. There can be **plateaus** and in some cases **peaks** in test-retest reliability and validity.
2. Asking more questions does **not** necessarily lead to better estimates of parameters.
3. A **decrease in reliability with more questions** asked is more likely when questions are asked in a **more efficient manner**, i.e., when convergence within a session is faster and/or when the use of heuristics increases at a slower rate.

Figure 1: Conceptual model* of preference measurement. Answering more questions leads to an increase in heuristic use. As heuristic use increases, we measure less of the participants true preference and more of the person and task specific heuristics

*We include a formal model in our working paper.

Experimental Evidence of Model

Reliability

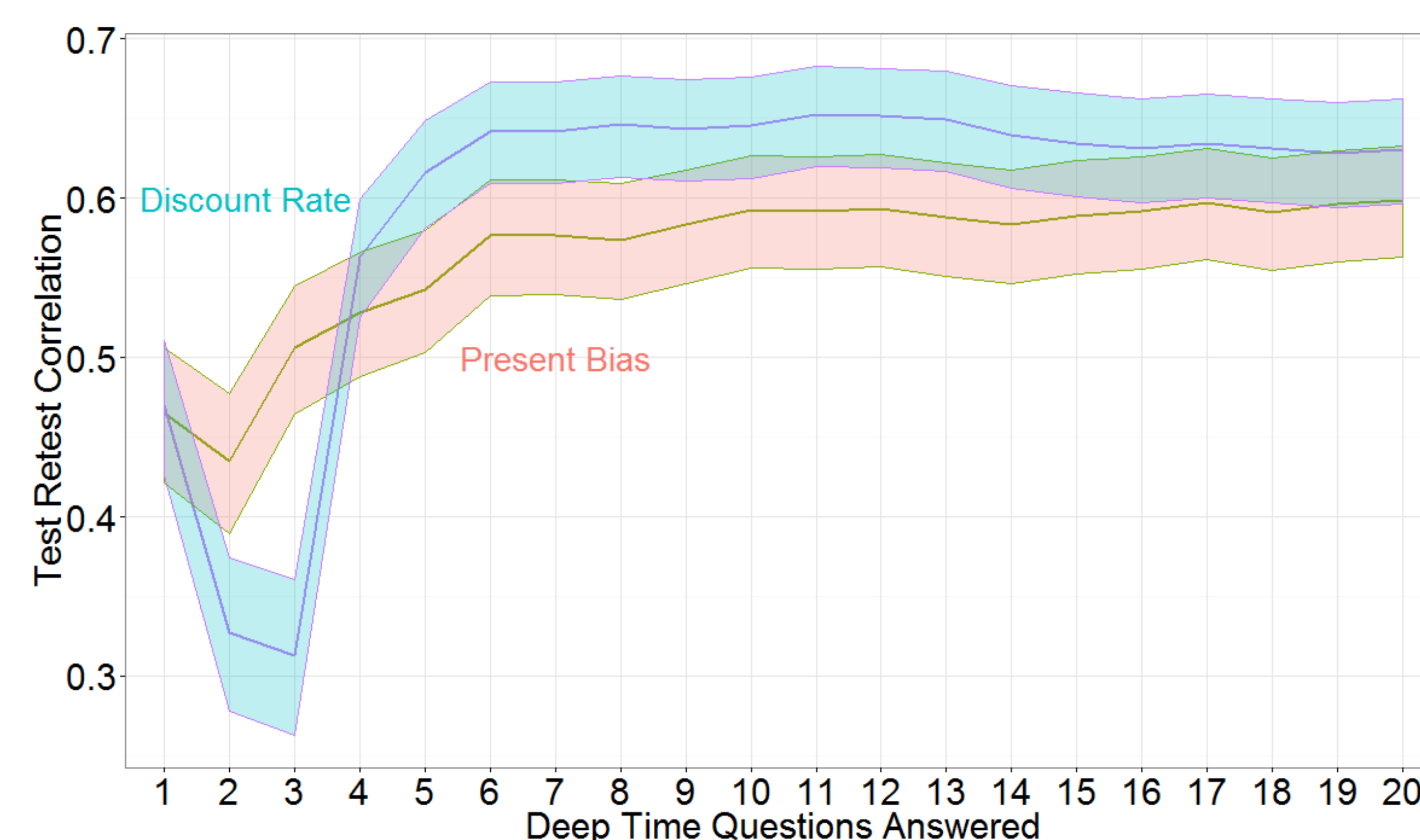


Figure 2: Test-retest reliability. 1308 Participants took DEEPTime 2 months apart and their Quasi-hyperbolic time discounting (QTD) parameters were estimated after each of 20 questions. We see a plateau in test-retest reliability. This plateau replicates in a sample of 444.

Note: DEEP Time is an adaptive choice task which measures Quasi-hyperbolic time discounting parameters. Bands in this and subsequent figures indicate 95% confidence intervals.

Concurrent Validity

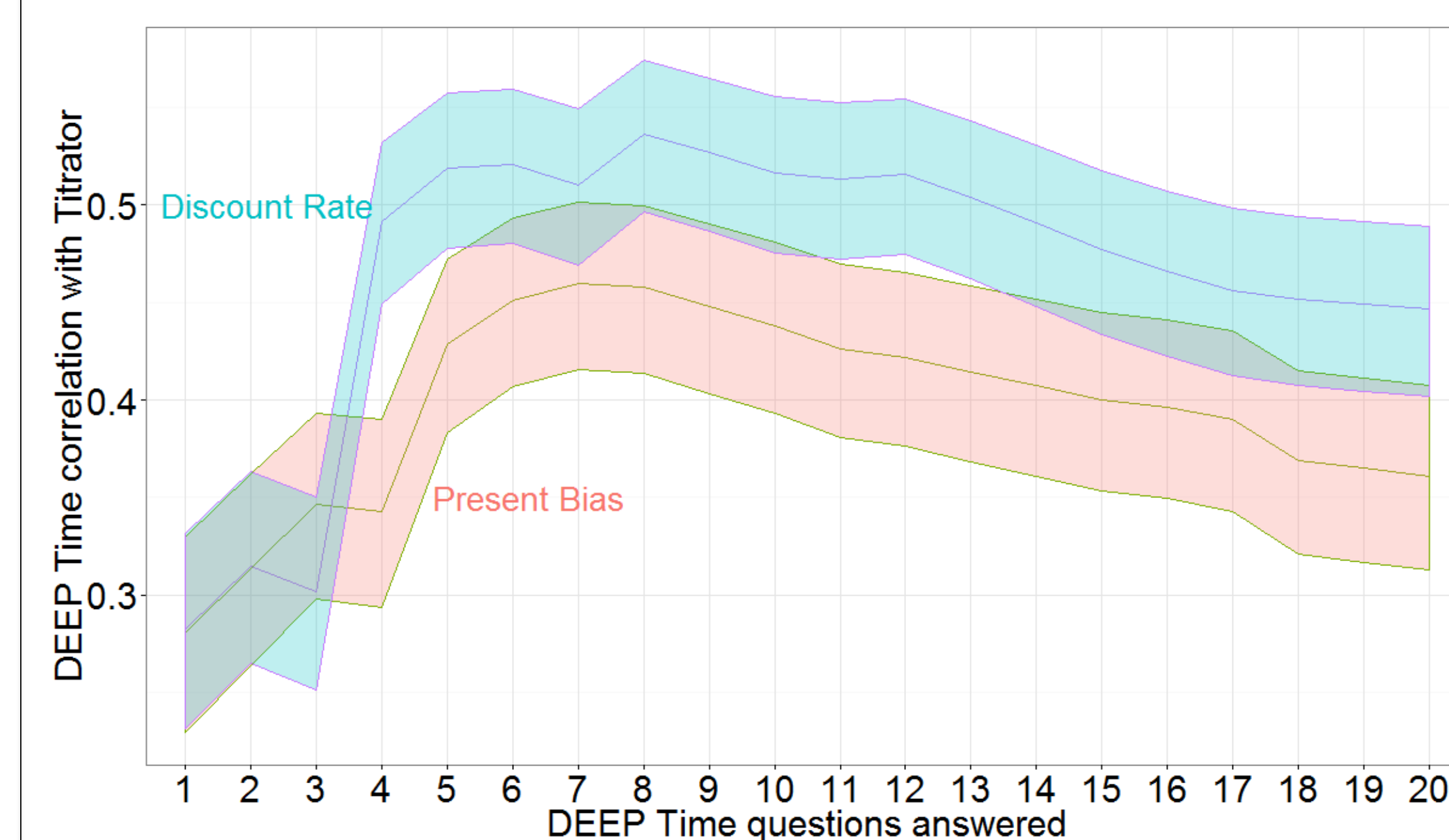


Figure 3: Concurrent Validity of time preferences. We see that the correlation between DEEP values and an external time preference titrator peak after 4 questions and decrease. A similar trend was found in other validity measurements.

As predicted by our model we see **plateaus in test-retest reliability and peaks in external validity.**

Conclusions

- Our generalized model of heuristic formation leads to novel predictions for preference measurement.
- We see a plateau in test-retest reliability.
- We see a marked decrease in concurrent validity with more questions asked.
- When measuring preferences, **less can be more.**

Future Directions

- How do context shifts alter heuristic use?
- Do estimates combined across contexts increase predictive validity?
- Do we see this pattern in other preference measurements (e.g., conjoint)?
- Testing the efficiency hypothesis.

References

- Toubia, O., Johnson, E., Evgeniou, T., & Delquié, P. (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science*, 59(3), 613-640.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.