

**Target article:** ‘The Generalizability Crisis’, by Tal Yarkoni.

**Title:** Is Formalism the Key to Resolving the Generalizability Crisis? An Experimental Economics Perspective.

**Author:** Zacharias Maniadis, University of Southampton, Department of Economics.

**Abstract:** I draw lessons from experimental economics. I argue that the lack of mathematical formalism cannot be usefully thought as the cause of the underappreciation of contextual and generalizability considerations. Instead, this lack is problematic because it hinders a clear relationship between theory and quantitative predictions. I also advocate a pragmatic policy-focused approach as a partial remedy to the generalisability problem.

**Formalism is Not Associated with a Strong Focus on Context.** Yarkoni (2021) correctly points out that the practice of using verbal models tested by statistical tools can be problematic. It then elaborates, focusing almost exclusively on the issue of context-dependence of human behavior and the overall complexity of the subject matter of social science, which renders generalizability of research findings difficult. I wish to contribute to the discussion from the perspective of a different behavioral discipline, experimental economics, which uses the mathematical language in its theory more frequently than psychology. I draw lessons from that discipline to show that the elaborated generalisability issues, while relevant, are not related directly to the lack of formalism.

In experimental economics, theories are predominantly mathematical, rather than verbal. Economics employs a set of principles, based on which, deductive models are constructed. These include preferences, beliefs, optimization, and equilibrium. Models and their associated properties are pitted against each other using data, while the formal rigor facilitates clear connection among models, underlying principles, and empirical methods. Prominent scholars have long regarded the assessment of competing models, not external validity, as the main focus of experiments (Smith, 1976; Plott, 1982). Schram (2005) argued that “... *external validity has received much more attention in psychology than in economics. To a large extent, psychological research is inductive and based on observed empirical regularities*”.

Camerer (2011) clearly explains why experimental economics has traditionally had a weaker concern for generalizability to real-life settings: “... *all empirical methods are trying to accumulate regularity about how behavior is generally influenced by individual characteristics, incentives, endowments, rules, norms, and other factors. A typical experiment therefore has no specific target for ‘external validity’ ...*”. According to this view – called the ‘scientific’ view – a theory-testing experiment helps choose between different theories and connects to our current understanding of the world.

Partly because of this specific methodological tradition, the issues that Yarkoni develops in the main text have not received major attention in experimental economics. As Loewenstein (1999) and Levitt and List (2007) argue, external validity or sampling concerns have not been given more focus relative to psychology – but see Exadaktylos et al. (2013) – and contextual variables are not regularly incorporated in models as Yarkoni envisions. Duflo (2017) argues: “... *details that we as economists might consider relatively uninteresting are in fact extraordinarily important in determining the final*

*impact of a policy or a regulation, while some of the theoretical issues we worry about most may not be that relevant”.*

A literature comparison indicates that experimental economists do not introduce and systematically vary contextual factors more frequently than psychologists (especially within a given study, as Yarkoni advocates). Because of their interest in general principles, economists focus more on the importance of homogenising important types of stimuli and removing context (Hertwig and Ortmann, 2001). However, Levitt and List (2007) argue that cross-situational consistency of behavior is lacking, which requires theories and methodologies to be addressed (for example, see Galizzi and Navarro-Martinez, 2019). Coinciding with a possible reproducibility crisis in science (see Ioannidis, 2005; Maniadis et al., 2014), theoretical interest in generalisability has recently increased (Zizzo, 2013; Kessler and Vesterlund, 2015; List, 2020).

To summarise the point: for experimental economics, it is not the case that the use of mathematical theories for decades was accompanied by a focus on the importance of heterogeneity of stimuli and other contextual factors. Instead, formal theory-testing is considered a domain where generalisability concerns should apply less. The problem of context-dependence in psychology may deserve to be addressed by careful statistical models and advanced experimental designs. However, the verbal representation of theories does not seem to be the culprit.

***Advantages of Formal Theory.*** I argue that the lack of formal theories in psychology is more problematic for another reason: it hampers clear theoretical predictions. In economics, formalism facilitates a relatively tight logical connection between theory and predictions. Accordingly, statistical research hypotheses follow theory naturally. Hence, it is more difficult to account – using ad hoc arguments – for experimental evidence inconsistent with a given theory. Muthukrishna and Henrich (2019) and Ortmann (2020) also advocate mathematical formalism to help us understand what a theory predicts and what it does not.

Contrary to the main connection made in Yarkoni (2021), a formal framework grounded on a set of overarching principles may facilitate knowledge accumulation not by allowing an arbitrary number of moderators to be considered, but by restricting the set of questions that are considered reasonable. This aspect of theory in experimental economics is now attracting some attention in psychology (Muthukrishna and Henrich, 2019). However, one needs to be cautious: while formalism makes excessive ad hoc theorising more difficult, it does not rule it out.

Experimental economics seems to fare better in terms of replicability (Camerer et al., 2016), and rigorous theory plays a role in this. However, this rigor mediates replicability primarily via some of the secondary channels mentioned in Yarkoni (2021): making riskier predictions and explicitly comparing competing theories. Predictions in economics tend to be much more quantitative and often estimation (rather than statistical hypothesis-testing alone) is the objective.

***A Pragmatic Approach.*** If the target is applicability to specific domains rather than theory-testing, another approach could be used. Randomised Controlled Trials in development and public economics examine the performance of interventions in natural environments. This methodological approach has been compared to that of plumbers, dentists or engineers (Roth, 2003, 2018; Duflo, 2017), and may be useful as a partial remedy to a possible ‘generalizability crisis’. Variability-enhancing designs that examine a high number of psychological factors may not always be pragmatic or feasible. Instead, in many cases of interest one could focus on specific policy domains and try to emulate them. A promising approach is assessing systematically whether the effect size of a given intervention is robust to the intervention being scaled-up as a full policy (Al-Ubaydli et al., 2017). Acknowledging the

importance of scalability in concrete policy domains could be a less ambitious – but potentially useful – approach for addressing a potential generalizability crisis.

## References

Al-Ubaydli, O., List, J. A., & Suskind, D. L. (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review*, 107(5), 282-86.

Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Available at SSRN 1977749.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433-1436.

Duflo, E. (2017). Richard T. Ely lecture: The economist as plumber. *American Economic Review*, 107(5), 1-26.

Exadaktylos, F., Espín, A. M., & Branas-Garza, P. (2013). Experimental subjects are not different. *Scientific reports*, 3(1), 1-6.

Galizzi, M. M., & Navarro-Martinez, D. (2019). On the external validity of social preference games: a systematic lab-field study. *Management Science*, 65(3), 976-1002.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists?. *Behavioral and Brain Sciences*, 24(3), 383-403.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

Kessler, J., & Vesterlund, L. (2015). The external validity of laboratory experiments: The misleading emphasis on quantitative effects (Vol. 18, pp. 392-405). Oxford, UK: Oxford University Press.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic perspectives*, 21(2), 153-174.

List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial (No. w27535). National Bureau of Economic Research.

Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *The Economic Journal*, 109(453), F25-F34.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1), 277-90.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 3(3), 221-229.

Ortmann, A. (2020). On the foundations of behavioural and experimental economics. Available at SSRN.

Plott, C. R. (1982). Industrial organization theory and experimental economics. *Journal of Economic Literature*, 20(4), 1485-1527.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4), 1341-1378.

Roth, A. E. (2018). Marketplaces, markets, and market design. *American Economic Review*, 108(7), 1609-58.

Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12(2), 225-237.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2), 274-279.

Yarkoni, T. (2021). The generalizability crisis. *Behavioral and Brain Sciences*, forthcoming.

Zizzo, D. J. (2013). Claims and confounds in economic experiments. *Journal of Economic Behavior & Organization*, 93, 186-195.