

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Fallucchi, Francesco; Marietta Leina, Andrea; Silva, Rui; Turocy, Theodore L.

Working Paper

A Comment on the "Strategic Complexity and the Value of Thinking by D. Gill and V. Prowse (2023)"

I4R Discussion Paper Series, No. 170

Provided in Cooperation with: The Institute for Replication (I4R)

Suggested Citation: Fallucchi, Francesco; Marietta Leina, Andrea; Silva, Rui; Turocy, Theodore L. (2024) : A Comment on the "Strategic Complexity and the Value of Thinking by D. Gill and V. Prowse (2023)", I4R Discussion Paper Series, No. 170, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/303908

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

INSTITUTE for **REPLICATION**

No. 170 I4R DISCUSSION PAPER SERIES

A Comment on the "Strategic Complexity and the Value of Thinking by D. Gill and V. Prowse (2023)"

Francesco Fallucchi Andrea Marietta Leina Rui Silva Theodore L. Turocy

October 2024



I4R DISCUSSION PAPER SERIES

I4R DP No. 170

A Comment on the "Strategic Complexity and the Value of Thinking by D. Gill and V. Prowse (2023)"

Francesco Fallucchi¹, Andrea Marietta Leina², Rui Silva², Theodore L. Turocy²

¹University of Bergamo/Italy ²University of East Anglia, Norwich/Great Britain

OCTOBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and metascientific work in the social sciences. Provided in cooperation with EconStor, a service of the <u>ZBW – Leibniz Information Centre for Economics</u>, and <u>RWI – Leibniz Institute for Economic Research</u>, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur University of Ottawa

Anna Dreber Stockholm School of Economics Jörg Ankel-Peters *RWI – Leibniz Institute for Economic Research*

E-Mail: joerg.peters@rwi-essen.de	Hohenzollernstraße 1-3	www.i4replication.org
RWI – Leibniz Institute for Economic Research	45128 Essen/Germany	

A comment on the "Strategic Complexity and the Value of Thinking by D. Gill and V. Prowse (2023)"*

Francesco Fallucchi Andrea Marietta Leina Rui Silva* Theodore L. Turocy

December 19, 2023

Abstract

Gill and Prowse (2023) study response times using a repeated p-beauty contest (p = 0.7). Looking at between-subject variation in response times, they found that subjects who think for longer, on average, win more rounds and choose lower numbers. When comparing average response times and level-k behavior, they observed that higher k types think for longer. In general, we are able to reproduce their findings, despite a minor coding error and some missing information. We test the robustness of their results by comparing average and median response times and choices, separating the sample into quick and slow respondents, including additional controls, and different estimation parameters. We do not find differences between choices between slow and quick respondents, somewhat contradicting their conclusions. Moreover, most subjects played faster as the game was repeated. The remaining results are robust to the inclusion of cohort effects and different parameter specifications in their regressions.

KEYWORDS: beauty contest; response times; level-k; strategic complexity

JEL CODES: C71; C92

^{*}Authors: Fallucchi: University of Bergamo. E-mail: francesco.fallucchi@unibg.it. Marietta Leina: University of East Anglia. E-mail: a.marietta-leina@uea.ac.uk. Silva: University of East Anglia. E-mail: r.constantino-da-silva@uea.ac.uk. Turocy: University of East Anglia. E-mail: t.turocy@uea.ac.uk. * - corresponding author. There is no conflict of interest with any of the original authors.

1 Introduction

Gill and Prowse (2023) use leverage response times from repeated strategic interactions in an attempt to measure the strategic complexity through the time people, on average, take in such an interaction. For that purpose, they use experimental data collected by Gill and Prowse (2016). In the respective experiment, 780 subjects were organized into 260 groups of three individuals. Each group played the p-beauty contest (with p = 0.7) for ten rounds with feedback but no rematching. The subjects had 90 seconds to make their choice in each round.

In the beauty contest with p < 1, subjects have an incentive to pick a number below the average choice. Iterative dominance reasoning, under strong epistemic assumptions, drives choices down to the lowest number in the action domain (the unique Nash equilibrium), especially when games are repeated. However, when a subject believes that others may select non-equilibrium numbers, there is an incentive to choose a number that may not be consistent with the Nash equilibrium of the game. Such a scenario is consistent with a heterogeneous distribution of subjects' reasoning levels (i.e., level-k behavior).

The authors initially look at the between-subject variation in response times, and find that subjects who think for longer, on average, win more rounds and choose lower numbers (closer to the equilibrium level). That result is consistent with the empirical evidence in Arad and Rubinstein (2012) and Branas-Garza et al. (2017). When comparing average response times with level-k behavior, they observe that higher level-k types think for longer, consistent with the evidence in Alós-Ferrer and Buckenmaier (2021) on one-shot p-beauty contests. They find considerable between-subject heterogeneity in how responsive thinking times are to changes in strategic complexity. They estimate a two-type mixture regression model and find that one type of subject varies their response times substantially with the strategic complexity of the situation that she faces, while the other type hardly varies their response times at all.

I4R DP No. 170

We are generally able to reproduce their findings. We found a minor coding error in their do-file, along with missing information on several variables, and the omission of the original (raw) data set on which the authors, later, fill in missing observations with imputed values. We further test the robustness of their results by comparing average and median response times and choices, separating the sample into quick and slow respondents, including additional controls, and testing different initial conditions on their mixture model. We do not find differences between the choices of slow and quick respondents, which somewhat contradicts their conclusions. Moreover, we find that the majority of subjects seem to have been playing faster as the game was repeated, an aspect masked when considering average behavior. The remaining results are, nonetheless, robust to the inclusion of cohort effects. We also do not have reasons to suspect that altering the initial conditions on their mixture model would change the main conclusions.

2 Reproducibility

The experimental data used in this paper was originally written in z-Tree with follow-up sessions in oTree. The data was mainly analyzed using Stata alongside a small routine in MatLab. We found a minor coding error in the do-file. Up to the section on "descriptives", the code runs without issues. At that point, the command "StataSubCode/DescIQPer.do Σ " produces an error due to the inclusion of the Σ symbol. The second issue, from a reproducibility standpoint, concerns the absence of information on the definition and labeling of a large number of variables in RTData_clean.dta. Examples include all the variables that start with mtResponse-TimeBC_XX. As a consequence, we struggled to understand which variables should be used when trying to reproduce the authors' results. On that note, we were not able to access the original raw data set without imputed values (which represent 0.3% of the responses on a set of twelve variables).

Regarding the MatLab code, we confirm that the routine runs without issues. The code was not only consistent but also easily verifiable, thanks to the reproducibility report and careful record-keeping by the authors.

3 Robustness

We turn our attention to the replication of the main findings. We discuss the robustness of their results regarding:

- 1. Response times and regression robustness;
- 2. Imputed values;
- 3. The fitting in the categorization of individuals (i.e., their mixture model approach).

3.1 Response times and regression models

When comparing the average and median response times (see Figure 1), we find substantial differences. By round 10, the median response time amounts to 8.5 seconds while the average is 25.6 seconds. This hints at the possibility that some subjects began playing faster at some point while others kept taking substantial amounts of time to make a choice. This simple comparison offers conflicting evidence with the possibility that response times were kept stable as argued in the original paper.

In Figure 2, we find a steadily clear decreasing trend in the average number chosen as the experiment is repeated. However, we find some additional puzzling evidence when considering the difference between average and median response time. More specifically, subjects who take more time, on average, to choose a number theoretically would pick lower numbers according to the original paper. But when we separate subjects into two groups, namely those who responded above and below the average response time in each round, we do not identify any significant differences between the average choices. Once again, this observation conflicts with the main conclusions in the paper. Regarding the regression analysis, the inclusion of cohort effects, which are missing from the original work, does not affect the statistical significance of the variables of interest. However, their magnitude changes by a substantial margin, in some specifications, as much as 75% (see Table 1). The model fitting also increases in a dramatic fashion. The results are also robust to the inclusion of round effects and the raven score (some other elements missing in the regressions in the original paper). In Table 2 [original paper], the inclusion of cohort effects does not affect the significance of both the Raven tests and the three personality factors.

3.2 Imputed values

As indicated by the authors in footnote 20 of the original paper, there are instances of missing data in the data collection, accounting for approximately 0.3% of the responses related to 12 items within the "Consideration of Future Consequences" scale. To address this issue, the authors substituted any missing responses with the average value derived from the non-missing responses for the corresponding questions. Regrettably, it is not possible to ascertain from the data set available online which specific answers were missing or to whom they belonged. The availability of such data would have opened the possibility of further tests on the robustness of the results.

3.3 Mixture model

Finite mixture models (Peel and McLachlan 2000) have become standard for categorizing individuals in experiments. However, the likelihood function of a finite mixture model usually shows irregularities such as multimodality (Lehmann and Casella 2006, Spiliopoulos et al. 2018). Our approach has been to test for this problem by changing the starting point of the fitting. However, the results appear to be robust to such different specifications.

I4R DP No. 170

4 Conclusion

In this brief note, we have commented on the reproducibility and replicability of the routines and findings in Gill and Prowse (2023). Overall, we are able to reproduce the code and routines used to produce the results in the original paper. However, in terms of replicability, we have some reserves concerning the conclusions of the paper. In particular, we find that the majority of participants, in fact, kept playing the game much faster as it was repeated (which is consistent with most learning patterns observed in such experiments). That aspect is masked when using averages given the tendency of some participants to take a long time to choose a number. Additionally, we are not able to find differences, on average, between the numbers chosen by quick and slow respondents. These findings offer some evidence against the possibility that longer response times may, eventually, be consistent with higher levels of reasoning.

References

- Alós-Ferrer, C. and Buckenmaier, J.: 2021, Cognitive sophistication and deliberation times, *Experimental Economics* 24, 558–592.
- Arad, A. and Rubinstein, A.: 2012, Multi-dimensional iterative reasoning in action: The case of the colonel blotto game, *Journal of Economic Behavior & Organization* 84(2), 571–585.
- Branas-Garza, P., Meloso, D. and Miller, L.: 2017, Strategic risk and response time across games, *International Journal of Game Theory* **46**, 511–523.
- Gill, D. and Prowse, V.: 2016, Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, *Journal of Political Economy* 124(6), 1619– 1676.
- Gill, D. and Prowse, V.: 2023, Strategic complexity and the value of thinking, The Economic Journal 133(650), 761–786.
- Lehmann, E. L. and Casella, G.: 2006, *Theory of point estimation*, Springer Science & Business Media.
- Peel, D. and McLachlan, G. J.: 2000, Robust mixture modelling using the t distribution, *Statistics and Computing* 10, 339–348.
- Spiliopoulos, L., Ortmann, A. and Zhang, L.: 2018, Complexity, attention, and choice in games under time constraints: A process analysis., *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44(10), 1609.

5 Figures



Figure 1: Extension of Figure 1 (a) in Gill and Prowse (2023)



Figure 2: Extension of Figure OA.1 in Gill and Prowse (2023)

6 Tables

	(1)	(2)	(3)	(4)
	miwinner_full	miprize_full	miLogprize_full	mibc_full
response time (test)	0.104^{***}	57.984***	0.258^{***}	-1.878***
	(0.024)	(13.552)	(0.060)	(0.662)
response time (paper)	0.060^{***}	38.118^{***}	0.160^{***}	-2.002**
	(0.019)	(9.027)	(0.042)	(0.872)
choices	7800	7800	7800	7800
$\operatorname{subjects}$	780	780	780	780
groups	260	260	260	260
R^2 (test)	0.299	0.032	0.107	0.695
R^2 (paper)	0.012	0.021	0.017	0.006

Table 1: Consistency checks: Tables 1 and 2 in Gill and Prowse (2023)

Notes: Standard errors in parentheses clustered at the cohort level. Rows in blue correspond to regressions with the inclusion of cohort fixed effects. Rows in red indicate the estimates in the paper. * (p<0.10), ** (p<0.05), *** (p<0.01)