

and results, we should too. The conclusion of the Science study on reproducibility in psychology [1] offers the following message:

Following this intensive effort to reproduce a sample of published psychological findings, how many of the effects can we confirm are true? Zero. And, how many of the effects can we confirm are false? Zero. Is this a limitation of the project design? No. It is the reality of doing science, even if it is not appreciated in daily practice.

Importantly, this is the reality of doing science *in general*, not only *social* science. We must not blind ourselves: our methods and protocols in HRI do not shelter us from the exact same problems experienced in other fields. Future researchers may well write the same kind of article about our field when they revisit today's literature on Human-Robot Interaction.

ACKNOWLEDGMENTS

This work has been partially supported by the EU H2020 L2TOR project (grant 688014), the EU FP7 DREAM project (grant 611391), the EU H2020 Marie Skłodowska-Curie Actions Innovative Training Networks project APRIL (grant 674868), and the EU H2020 Marie Skłodowska-Curie Actions project DoRoThy (grant 657227). All authors have contributed equally to the experimental design, execution, data analyses and writing.

REFERENCES

- [1] A.A. Aarts, C.J. Anderson, J. Anderson, M.A.L.M. van Assen, P.R. Attridge, et al. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015). <https://doi.org/10.1126/science.aac4716>
- [2] F.H. Allport. 1924. *Social Psychology*. Houghton Mifflin Company, Chapter Response to social stimulation in the group, 260–291.
- [3] M. Baker. 2016. 1500 scientists lift the lid on reproducibility. *Nature News* 533, 7604 (2016), 452. <https://doi.org/10.1038/533452a>
- [4] C. Bartneck. 2011. The End of the Beginning: A Reflection on the First Five Years of the HRI Conference. *Scientometrics* 86, 2 (2011), 487–504.
- [5] C. Bartneck. 2017. Reviewers' scores do not predict impact - Bibliometric Analysis of the Proceedings of the Human-Robot Interaction Conference. *Scientometrics* 110, 1 (2017), 179–194. <https://doi.org/10.1007/s11192-016-2176-y>
- [6] P. Baxter, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme. 2016. From characterising three years of HRI to methodology and reporting recommendations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 391–398.
- [7] C.F. Bond. 1982. Social Facilitation: A self-presentational view. *Journal of Personality and Social Psychology* 42, 6 (1982), 1042–1050.
- [8] C.F. Bond and L.J. Titus. 1983. Social facilitation: a meta-analysis of 241 studies. *Psychological bulletin* 94, 2 (1983), 265–292. <https://doi.org/10.1037/0033-2909.94.2.265>
- [9] J. Cohen. 1977. *Statistical power analysis for the behavioral sciences*. Academic Press.
- [10] N.B. Cottrell, R.H. Rittle, and D.L. Wack. 1967. The presence of an audience and list type (competitional or noncompetitional) as joint determinants of performance in paired associates learning. *Journal of Personality* 35 (1967), 425–434.
- [11] W.D. Criddle. [n. d.]. The physical presence of other individuals as factor in social facilitation. *Psychonomic Science* 22, 4 ([n. d.]), 229–230.
- [12] J.M. Feinberg and J.R. Aiello. 2010. The Effect of Challenge and Threat Appraisals Under Evaluative Presence. *Journal of Applied Social Psychology* 40, 8 (2010), 2071–2104.
- [13] T.R. Fosgaard, L.G. Hansen, and M. Piovesan. 2013. Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior and Organization* 93 (2013), 279–284. <https://doi.org/10.1016/j.jebo.2013.03.027>
- [14] V.J. Ganzer. 1968. Effects of audience presence and test anxiety on learning and retention in a serial learning situation. *Journal of Personality and Social Psychology* 8, 2 (Pt. 1) (1968), 194–199.
- [15] W.L. Gardner and M.L. Knowles. 2008. Love Makes You Real: Favorite Television Characters Are Perceived as "Real" in a Social Facilitation Paradigm. *Social Cognition* 26, 2 (2008), 156–168. <https://doi.org/10.1521/soco.2008.26.2.156>
- [16] R.G. Geen. 1973. Effects of being observed on short- and long-term recall. *Journal of Experimental Psychology* 100 (1973), 395–398.
- [17] B. Guerin. 1983. Social Facilitation and social monitoring: a test of three models. *British Journal of Social Psychology* 22 (1983), 203–214. <https://doi.org/10.1111/j.2044-8309.1983.tb00585.x>
- [18] B. Guerin. 1989. Reducing evaluation effects in mere presence. *The Journal of Social Psychology* 129, 2 (1989), 183–190.
- [19] J.P. Hill and R.A. Kochendorfer. 1969. Knowledge of peer success and risk of detection as determinants of cheating. *Developmental Psychology* 1, 3 (1969), 231–238. <https://doi.org/10.1037/h0027330>
- [20] L.K. John, G. Loewenstein, and D. Prelec. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science* 23, 5 (2012), 524–532.
- [21] S.R.G. Jones. 1992. Was There a Hawthorne Effect? *Amer. J. Sociology* 98, 3 (1992), 451–468.
- [22] J.P. Lombardo and J.F. Catalano. 1975. The effect of failure and the nature of the audience on performance of a complex motor task. *Journal of Motor Behavior* 7 (1975), 29–35.
- [23] R.J. McCaffrey, J.M. Fisher, B.A. Gold, and J.K. Lynch. 1996. Presence of third parties during neuropsychological evaluations: Who is evaluating whom? *The Clinical Neuropsychologist* 10, 4 (1996), 435–449. <https://doi.org/10.1080/13854049608406704>
- [24] J. McCambridge, J. Witton, and D.R. Elbourne. 2014. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology* 67, 3 (2014), 267–277.
- [25] F.G. Miller, M.E. Hurkman, J.B. Robinson, and R.A. Feinberg. 1979. Status and evaluation potential in the social facilitation and impairment of task performance. *Personality and Social Psychology Bulletin* 5 (1979), 381–385.
- [26] D.S. Nagin and G. Pogarsky. 2003. An Experimental Investigation of Deterrence: Cheating, Self-Serving Bias, and Impulsivity. *Criminology* 41, 1 (2003), 167–194. <https://doi.org/10.1111/j.1745-9125.2003.tb00985.x>
- [27] N. Riether, F. Hegel, B. Wrede, and G. Horstmann. 2012. Social facilitation with social robots?. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12*. 41–47. <https://doi.org/10.1145/2157689.2157697>
- [28] F.J. Roethlisberger, W.J. Dickson, H.A. Wright, and Western Electric Company. 1939. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago*. Harvard University Press.
- [29] M.S. Rosenberg. 2005. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution* 59, 2 (2005), 464–468.
- [30] R. Rosenthal. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86, 3 (1979), 638–641.
- [31] H.R. Rothstein, A.J. Sutton, and M. Borenstein. 2006. *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.
- [32] P. Schermerhorn, M. Scheutz, and C.R. Crowell. 2008. Robot Social Presence and Gender: Do Females View Robots Differently than Males? *ACM/IEEE International Conference on Human-Robot Interaction* (2008), 263–270. <https://doi.org/10.1145/1349822.1349857>
- [33] J.P. Simmons, L.D. Nelson, and U. Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science* 22, 11 (2011), 1359–1366.
- [34] W. Stroebe. 2012. The truth about Triplett (1898), but nobody seems to care. *Perspectives on Psychological Science* 7, 1 (2012), 54–57.
- [35] M.J. Strube, M.E. Miles, and W.H. Finch. 1981. The social facilitation of a simple task: Field tests of alternative explanations. *Personality and Social Psychology Bulletin* 7 (1981), 701–707.
- [36] D.J. Terry and M. Kearnes. 1993. Effects of an audience on the task performance of subjects with high and low self-esteem. *Personality and Individual Differences* 15, 2 (1993), 137–145.
- [37] N. Triplett. 1898. The dynamogenic factors in pacemaking and competition. *American Journal of Psychology* 9, 4 (1898), 507–533.
- [38] L. Uziel. 2007. Individual differences in the social facilitation effect: A review and meta-analysis. *Journal of Research in Personality* 41, 3 (2007), 579–601. <https://doi.org/10.1016/j.jrjp.2006.06.008>
- [39] F.T. Vitro and L.A. Schoer. 1972. The effects of probability of test success, test importance, and risk of detection on the incidence of cheating. *Journal of School Psychology* 10, 3 (1972), 269–277. [https://doi.org/10.1016/0022-4405\(72\)90062-3](https://doi.org/10.1016/0022-4405(72)90062-3)
- [40] K.D. Vohs and W. Schooler. 2008. The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating. *Psychological Science* 19, 1 (2008), 49–54. <https://doi.org/10.1111/j.1467-9280.2008.02045.x>
- [41] I. Wechsung, P. Ehrenbrink, R. Schleicher, and S. Möller. 2014. Investigating the Social Facilitation Effect in Human-Robot Interaction. In *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems into Practice*, J. Mariani, S. Rosset, M. Garnier-Rizet, and L. Devillers (Eds.). Springer, New York, 167–177.
- [42] R.B. Zajonc. 1965. Social facilitation. *Science* 149, 3681 (1965), 269–274.