

Social Psychological Bulletin

Psychologia Społeczna

How to Make Psychology a Genuine Science of Behavior: Comment on Dolinski's Thoughtful Paper

Klaus Fiedler¹

¹ *University of Heidelberg, Heidelberg, Germany*

Corresponding author: Klaus Fiedler (University of Heidelberg, Hauptstraße 47-51, 69117 Heidelberg, Germany.
E-mail: kf@psychologie.uni-heidelberg.de)

Received: 21 January 2018 • **Published:** 29 May 2018

Citation: Fiedler, K. (2018). How to make psychology a genuine science of behavior: Comment on Dolinski's thoughtful paper. *Social Psychological Bulletin*, 13(2), Article e26079. <https://doi.org/10.5964/spb.v13i2.26079>

Abstract

In this comment to Doliński's (2018, this issue) challenging paper, I express my agreement with his basic ideas and with his concerns about the alienation of social psychology. However, I also present some critical thoughts that amount to a slightly different diagnosis of the present situation. Rather than concluding that our discipline has ceased to study real behaviors, I provide positive counter-examples of substantial behavioral science and argue that the major problem is not to distinguish between measures of "real" and "non-real" behaviors. The problem core, rather, lies in the widespread tendency to mistake statistical and technical indices (latencies, model parameters, fMRI indices, etc.) for measures of meaningful behavior. When technical means become ends in themselves, Doliński's metaphor applies that "the tail wags the dog".

Keywords

verbal behavior; compliance; content validity; symbolic behavior

The question raised by Dariusz Doliński (2018, this issue) strikes me as challenging, important and timely: "Is psychology still a science of behavior?" An analysis of the contents of an entire volume of the flagship "Journal of Personality and Social Psychology" (*JPSP*) seems to disclose an alarming state of affairs: "Out of a total of 290 studies presented in the volume of *JPSP* under analysis, a mere 18 [...] addressed behaviors" (p. 7), as distinguished from finger movements required for mouse clicks or survey responses. The rate of genuine behavioral studies that Baumeister, Vohs, and Funder (2007) had es-



estimated to be 80% in 1976 and still over 10% in 2006, the prevalence of truly behavioral studies has, according to Doliński (2018, this issue), further shrunk to a negligible level. Given the major role assigned to behavioral science in dealing with many challenges of the 21st century – such as migration, inequality, new media, terrorism, and aging – this strikes me as a memorable and hardly credible picture of our discipline. Diagnosing and overcoming this state of affairs constitutes a prominent developmental task for all behavioral scientists.

Some Critical Reflections Triggered by Doliński's Paper

Although I am convinced that Doliński's critique is justified and that his paper has the potential to initiate a valuable discourse and an overdue contemplation process, I would like to raise some critical remarks. First, if we complain about the paucity of research assessing actual behavior, it would be essential to precisely define the distinction between genuine behavior and mere self-reports, mouse clicks, and as-if behaviors. In this regard, I would like to see the coding criteria for the 290 studies summarized in Doliński's Table 1. And, I would like to discuss, more generally, what dependent measures qualify as appropriate, real behavior.

Is Verbal Behavior Less Real Than Motor Behavior?

For instance, I emphatically contest that verbal behavior is less real, less consequential, or less representative of social life than motor behavior, manifest aggression, or expressed emotions. In modern civilized cultures, almost all social and political decisions and actions – in legislative, executive, and juridical context – is mediated by language-based deliberation, constrained by codified law, and based on education that relies on language. Verbal behavior engulfs advertising, flirting, psychotherapy, and instruction, and verbal intelligence is no doubt more influential than fists or kick-boxing. Clicking with a computer mouse, too, has become a committal and consequential means of social action and interaction, to make friends on social media, to sign contracts, or to explore and gather information about our ultimate objects of interest.

Secondly, however constructive and well-motivated Doliński's paper may be, I believe that more influence can be attained by highlighting existing positive examples than by complaining about examples of misdirected research. I would indeed argue that – regardless of the present content analysis of *JPSP* – a good deal of compelling behavioral research does exist, at least according to my definition of “real behavior”. To list but a few positive examples, experiments on evaluative conditioning evoke real evaluative reactions (cf. special issue 2017 in *Social Cognition*), research on stereotype threat refers to actual performance impairment (Régner, Steele, & Huguet, 2014), as does ego-depletion research (Hagger, Wood, Stiff, & Chatzisarantis, 2010), serial reproduction involves real communication chains (Lyons & Kashima, 2006), group decision making speaks immediately to democratic societies (Leach, 2016), and the APS journal *Psychological Science in the Public*

Interest covers manifold streams of impressive behavioral research, such as women in academic settings (Ceci, Ginther, Kahn, & Williams, 2014), communication of medical risk (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007), and fairness in tax paying and other economic behaviors (Muehlbacher et al., 2008).

“When the Tail Wags the Dog”

While it is easy to provide more positive examples – which deserve to be imitated and to be given more attention than negative states of affairs – this should not prevent us from admitting that Dariusz Doliński’s critique is up to the point and must not be ignored. In my attempt to diagnose the origin of the disease, I came to conclude that a main reason for alienation and emptiness of much of the current research lies in the perverted relation of means and ends. “The tail wags the dog” (p. 9), that is, the methods, paradigm labels, and fashionable phrases, which are nothing but means for conducting substantial tests of theoretical and practical problems, have become ends and research goals in their own right. Reduced recognition thresholds for aggression-related words are treated as valid measures of (implicit) aggression, and a whole variety of chronological measures are considered manifestations of implicit stereotypes, attitudes, desires or even implicit associations. It ought to be clear that such symptoms of lexical accessibility are neither necessary nor sufficient ingredients of real attitudes, that nobody has ever explicated (and hardly anybody will ever be able to explicate) the difference between implicit and explicit associations. The journal readers, reviewers, and editors who constitute the scientific community do not mind mistaking a superficial attention check for a genuine manipulation check supposed to validate an effective change in an independent variable. A lousy single-item introspective rating is readily accepted as a viable dependent variable if only the study is pre-registered. A shallow hypothesis test becomes excellent science to be published in leading journals as soon as it is re-labelled as model testing. Or, an otherwise boring piece of correlational research seems to meet high standards as soon as a confound of the dependent variable is entered into a mediation test and (unsurprisingly) shown to reduce the partial correlation between independent and dependent variable (Fiedler, Harris, & Schott, 2018).

The common denominator and major reason for such alienation lie – in my opinion – in the scientific community’s failure to engage in critical examination of mainstream products and fashionable ideas. The chief determinant of the current *Zeitgeist* is apparently compliance, conceived as an uncritical acceptance of norms and (majority) rules, the justification of which is hardly ever called into question. It is as if the lesson gained from Hannah Arendt’s (1963) book on the banality of the evil half a century ago – that we have no right to be just obedient and to comply with questionable norms – was in vain. To be sure, the scientific community does have the mental equipment to understand, and was actually sensitized repeatedly to the fact that a fast mouse click may not be a prejudiced attitude, that a good model fit tells us little about an underlying mechanism (Roberts & Pashler, 2000), that blood flow in the brain must not be confused with manifest behavior, and that

an exact p value from a null-hypothesis significance test does NOT have much evidential value. Would the scientific community – that is, the editors, reviewers, journal readers, students, and WE ALL – engage in the sort of critical examination that our beloved scientific discipline calls for, it would hardly be possible to sell us surrogates for real behavior, tails that wag the dog, or mediation mimicry for genuine mechanisms.

How to Overcome the Current Stage of Alienated (Social) Psychology

Any future attempts to overcome the dissatisfying state identified by Dariusz Doliński must in my opinion start from the same self-attribution: We all (i.e., the scientific community) must jointly analyze, understand, and correct for the problem. No doubt, the ultimate gatekeepers are the journal editors and editorial boards. Their influence and their responsibility cannot be overestimated. Being myself an Associate Editor of the *Journal of Personality and Social Psychology*, which has been the target of critical content analysis, I have decided to carry Doliński's memorable point to the next editorial meeting and to start a discourse among editors. And, I will share it with other editors of leading journals, who function as trendsetters in social psychology. We all should try to contribute to a serious attempt to overcome shallow science. We can devote edited volumes and special issues or special-issue sections to the investigation of "real behavior", we can organize conferences or summer schools on this exciting topic, and we can involve the reviewers and agents of our national funding schemes.

Most importantly, though, we should mind our own business and take to heart the ideal that good research must be based on sound measures of behavior. And here we return to the crucial question of what qualifies as genuine behavior. What paradigms, what experimental tasks, what behavioral measures, and what motivating research contexts should we employ to make psychological findings diagnostic of such issues as aggression and altruism, defection and cooperation, discrimination and integration, health, problem solving, fairness, problem solving, risk management and adaptive behavior?

My tentative answer to this crucial question leads to the following considerations. First, it goes without saying that it is ethically and psychologically impossible to induce motives and observe or measure behavior at the same level of intensity and existential significance as in real life. We cannot directly study murder, suicide, car accidents, rape and sexual assault, love, and marriage, serious risk taking, property loss, jealousy, or depression under controlled conditions.

However, secondly, we can nevertheless conduct studies that allow us to make reasonable inferences about such intense and consequential behaviors. A variety of experimental paradigms focus on manifest, content-valid behaviors: we can study problem solving, fairness in reward allocation, politeness in discussions, social hypothesis testing, ingratiation, politeness and self-presentation in natural interactions, lie production and lie detection under non-trivial conditions. Although the payoffs and personal consequences that are at

stake in such experiments are clearly smaller than in most existential real life settings, we can build on the content validity of such genuine behavioral measures.

Thirdly, merely symbolic behaviors that do not meet the criterion of content validity can be highly relevant to understanding real behavior, as illustrated in game theory. Even when we let participants play the prisoner's dilemma or coordination games with play money, using mouse clicks as a response mode, surrounded by arbitrarily constructed cover stories, the behavioral strategies and adaptive reactions in such artificial games can foster deep insights to be validated in more naturalistic settings. The choices observed in such games constitute real choices. For instance, investigating in a typical time-discounting paradigm the amount of (play) money that people are ready to sacrifice if they don't have to wait for an outcome can be enlightening and scientifically important, even though the outcome is only a few Euros or only symbolic.

As game theory is often represented in formal notation, it can be used, forth, to explain how formal, mathematical reasoning can make strong and substantial contributions to behavioral science. Krueger's (2013) argument that projection (i.e., the tendency to believe that many other people behave as we do) can lead to trust and pro-social behavior in coordination games strikes me as an inspiring example. Another example would be the sampling-theoretical proof that outgroup homogeneity can be derived from the mathematical truism that, given larger samples of observations for ingroups than outgroups, the experienced variation of outgroups is in fact reduced, in the absence of any cognitive or motivational bias.

So, if all these cases of content-valid measures, symbolic behaviors, and even purely theoretical and mathematical arguments can be justified as highly relevant to the analysis of behavior, what other research may Doliński refer to? Let me try to be very explicit about what I consider to be the core of his critique. I believe the problem does not lie in the selection of the wrong variables but in the unwarranted theoretical and practical meaning given to the variables being manipulated and measured in psychological science. Rather than excluding verbal measures, or symbolic behavior, or mouse clicks or any other response modes as essentially irrelevant or unreal, I believe the essential problem lies in Doliński's notion of the tail wagging the dog. That is, rather than demonstrating how projection induces trust and coordination (measured in terms of play money, which would be fine), the focus is only on fitting a model, making a technical (and often logically unwarranted) point for a mediation model, an arbitrary allusion to what can be considered rational, or on some allegedly implicit measure of cooperation.

Even when choices, preferences, verbal aggressions, or punishments are studied with play money, using trivial consumer products, or unknown target persons, they are still genuine choices, preferences and verbal actions (the generality of which remains to be tested). However, when response latencies are offered as attitudes, an averaging rule as a cognitive mechanism, or increased error rates under time pressure as diagnostic indices of "system 2", then we fall prey to category mistakes. Neither a bias parameter in signal detection nor a priming effect nor eye-tracking indices, ERP or even blood-flow (fMRI) indices, are suf-

ficient to measure behavior proper. And, it is insufficient, and close to self-deception, to believe that one has measured X simply because some authority has attached the label X (e.g., “heuristic”; “automatic”; “implicit association”) to the dependent variable, or simply because participants have been prompted to rate their own “X”. In these cases, when para-behavioral measures, technical parameters, and sonorous verbal labels referring to nothing but phantoms become ends in and of themselves, then the tail indeed wags the dog. This is in my view the real problem – a problem of referentiality and theoretical interpretation rather than a matter of selecting privileged measures of “real behaviors”.

Funding

The author has no funding to report.

Competing Interests

The author has declared that no competing interests exist.

Acknowledgements

The author has no support to report.

References

- Ahrendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. London, United Kingdom: Faber & Faber.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4), 396–403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15(3), 75–141. <https://doi.org/10.1177/1529100614541236>
- Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, 13(2), Article e25025. <https://doi.org/10.5964/spb.v13i2.25025>
- Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests – An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, 75, 95–102. <https://doi.org/10.1016/j.jesp.2017.11.008>
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53–96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>

- Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495–525. <https://doi.org/10.1037/a0019486>
- Krueger, J. I. (2013). Social projection as a source of cooperation. *Current Directions in Psychological Science*, *22*(4), 289–294. <https://doi.org/10.1177/0963721413481352>
- Leach, D. K. (2016). When freedom is not an endless meeting: A new look at efficiency in consensus-based decision making. *The Sociological Quarterly*, *57*(1), 36–70. <https://doi.org/10.1111/tsq.12137>
- Lyons, A., & Kashima, Y. (2006). Maintaining stereotypes in communication: Investigating memory biases and coherence-seeking in storytelling. *Asian Journal of Social Psychology*, *9*(1), 59–71. <https://doi.org/10.1111/j.1467-839X.2006.00184.x>
- Muehlbacher, S., Kirchler, E., Hoelzl, E., Ashby, J., Berti, C., Job, J., . . . Waldherr, K. (2008). Hard-earned income and tax compliance: A survey in eight nations. *European Psychologist*, *13*(4), 298–304. <https://doi.org/10.1027/1016-9040.13.4.298>
- Régner, I., Steele, J. R., & Huguet, P. (2014). Stereotype threat in children: Past and present. *Revue Internationale de Psychologie Sociale*, *27*(3-4), 5–12.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. <https://doi.org/10.1037/0033-295X.107.2.358>