Cognitive Processing (2019) 20:273–275
https://doi.org/10.1007/s10339-019-00913-2

OPINION PAPERS AND COMMENTARIES

Empirical content as a criterion for evaluating models

Marc Jekel[1]

Received: 22 February 2019 / Accepted: 11 March 2019 / Published online: 20 March 2019
© Marta Olivetti Belardinelli and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Hypotheses derived from models can be tested in an empirical study: If the model reliably fails to predict behavior, it can be dismissed or modified. Models can also be evaluated before data are collected: More useful models have a high level of empirical content (Popper in Logik der Forschung, Mohr Siebeck, Tübingen, 1934), i.e., they make precise predictions (degree of precision) for many events (level of universality). I apply these criteria to reflect on some critical aspects of Kirsch's (Cognit Process, 2019. https://doi.org/10.1007/s10339-019-00904-3) unifying computational model of decision making.

Keywords Model evaluation · Empirical content · Universality · Precision · Falsification · Theory of science

Alexandra Kirsch (2019) submitted her article on the unifying computational model of decision making. Computational models force researchers to specify the variables of a theory and the functional relation between variables and predicted behavior. Models with such a high level of specification have many advantages over typically less well-specified non-formalized "verbal" models (Farrell and Lewandowsky 2010) that are more common in psychology (Fiske 2004). They can be used in "thought experiments [that are] prosthetically regulated by computers" (Dennet 1981, p. 117)[1] to, for example, derive hypotheses that can be tested empirically. A unifying model can integrate complementing theories, make relations between empirical phenomena visible (Fiedler 2004; Van Lange et al. 2012), and spark new research. The model by Kirsch (2019) has these qualities of computational models, and I agree with Ross (2019) that its value will be finally measured by the extent in which it initiates research. Nonetheless, part of the role of a reviewer of an article is to evaluate its content and I had some concerns about the model, arguing from a Popperian perspective that may be useful to consider or debate (Ross 2019) when evaluating and developing models.

Criteria for evaluating models before and after data are collected

Experiments and correlational studies in the laboratory and in the field allow testing whether hypotheses derived from models[2] are in line with the predicted behavior. If studies allow for strong inferences (Platt 1964; Roberts and Pashler 2000) and a model reliably fails to make correct predictions, the model either needs to be adjusted or dismissed. If the model makes correct predictions only for some participants or only under some conditions, the generality of the model needs to be reconsidered. If the model makes risky predictions that, in principle, could turn out to be false, but nonetheless predicts behavior accurately, its degree of corroboration increases (Popper 1934).

Generally less common, at least in psychology, is the use of criteria for evaluating theories *before* data are collected (i.e., a priori) such as the empirical content of a theory (Popper 1934). A theory has a high level of empirical content if it achieves a high level of universality and a high degree of precision.

A theory is more universal if it applies to more observable events. More formally, the level of universality of a theory can be assessed by the extent to which the "if" statements in a hypothesis restrict the number of events the theory can be applied. For example, the hypothesis "*if* a child is frustrated, *then* it reacts aggressively" cannot be applied to

[1] I thank the reviewer who made me aware of this work.

[2] I use the terms theory and model interchangeably but see, in contrast, Thagard (2012, Chapter 1) for a differentiation.

Marc Jekel
marc.jekel@uni-koeln.de

[1] Social Cognition Center Cologne, University of Cologne, Richard-Strauss-Straße 2, 50931 Cologne, Germany

predict behavior of frustrated adults. As another example, the theory of cognitive dissonance (Festinger 1957) cannot be applied to predict (lack of) attitude change in cases where cognition and behavior is coherent (Higgins 2004).

A theory is more precise if the "then" statements in a hypothesis are more restrictive in the sense that most behavior that could be potentially observed in a study contradicts the theory. More formally, the degree of precision of a theory can be quantified by the extent of prediction space that contradicts the theory (see Figure 1, Roberts and Pashler 2000). In the example of the relation between frustration and aggression, the theory can only be falsified by a negative or no relation between the two variables but is consistent with any positive relation whatsoever between them. A more precise theory, however, could predict a linear relation between both variables and thus define the exact functional relation between the extent of frustration and aggressive behavior. Similarly, a formal implementation of cognitive dissonance (Shultz and Lepper 1996) might be more specific regarding the functional relation between dissonance and attitude change.[3]

The degree of precision is closely related to model flexibility (Pitt and Myung 2000; Roberts and Pashler 2000), that is, models that predict almost any behavior that could be observed are overly flexible and thus lack precision. Statistical methods have been developed to punish for model flexibility when evaluating the fit of a model (Myung and Pitt 1997; Myung et al. 2006; Pitt et al. 2002). Model flexibility may also be restricted a priori, for example, by using priors on model parameters (Vanpaemel and Lee 2012) and thus priors on predicted behavior.

The method of computational modeling may naturally help to foster better reasoning about models in general (Farrell and Lewandowsky 2010; Marewski and Olsson 2009) and also in accordance with criteria of universality and precision (cf. Klein 2014). A computational model typically consists of functions written in some programming language that have input arguments (i.e., variables of the if statement in the hypothesis). In the body of the functions, a computational model defines the functional structure by which the arguments are combined to produce the predicted behavior (i.e., variables of the then statement in the hypothesis). The flexibility of a computational model and thus its degree of precision can be easily evaluated before data are collected by plotting all model predictions that are consistent with the theory in the prediction space.

Models with a high level of empirical content are a priori more useful in practice: They can be applied in many situations (criterion of universality), and they make more informative predictions (criterion of precision). Models with a high level of empirical content are also more useful for theory development: They can be tested and thereby falsified in many situations. They also make more specific and thus more risky predictions that can be falsified more easily. Incorrect models can therefore be detected and dismissed (or modified) more easily if they have a high level of empirical content.

Some critical aspects of Kirsch's (2019) unifying computational model of decision making

The unifying computational model of decision making (Kirsch 2019) defines all the necessary consecutive operations in the decision process in reasoning tasks such as getting all the alternatives and all the cues from memory or the environment, ordering and aggregating cues, stopping aggregation based on some criterion of acceptability, deciding between options, or adding more options. This rather loose framework can be filled with if/then statements (Kirsch 2019, Figure 2, p. 3) and computations for each operation to model rule-based decision making. Parameters in the model can instantiate specific computations such as computational steps of heuristics from the adaptive toolbox (Gigerenzer and Todd 1999).

This "generalized model" (Kirsch 2019, p. 2) can reproduce heuristics and thus should be able to predict human decision making well. The model can also instantiate other algorithms from, for example, artificial intelligence that lead to good performance in a decision task but are not necessarily plausible candidates for modeling human decision making. Thus, the model may be used not only for predicting human behavior (i.e., descriptive purpose) but also for solving decision tasks efficiently (i.e., prescriptive purpose) as also demonstrated in a case study in the article.

The unifying computational model of decision making achieves a high level of universality. In the current version, its application is not restricted to specific groups of persons. The decision models that can be instantiated in the generalized model could also stem from diverse domains such as probabilistic reasoning (e.g., Gigerenzer and Goldstein 1996), risky choice (e.g., Kahneman and Tversky 1979), or preferential choice (e.g., Krajbich et al. 2010). Thus, the application of the model is also not restricted to a specific domain.

The unifying computational model of decision making, however, does not (in my view) achieve a high degree of precision in its current version which might, of course, change in an updated version. For example, the article focuses on the operation of information aggregation for demonstrating its generality. For aggregation, the "model allows for *any* [emphasis added] aggregation mechanism in the context of

[3] See also Glöckner and Betsch (2011) for a recent application of these criteria for evaluating models in judgment and decision making.

a full, iterative decision procedure" (Kirsch 2019, p. 2). If any aggregation mechanism is possible, it is unclear how the model decides which mechanism to choose for a specific application. This crucial aspect of the model is not very well specified since "[t]he parameters, however, still have to be found for each application in turn" (Kirsch 2019, p. 2). This is especially worrisome because some of the models listed in the article that can be instantiated in the unifying computational model of decision making are competitor models, i.e., they can make differing predictions (Glöckner et al. 2014; Rieskamp and Otto 2006). After all, if it were the case that *every* prediction was possible, then the model could not be falsified and thus the model would not contain any empirical content.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

Human and animal rights statements This article does not contain any studies with human participants or animals performed by the author.

References

Dennet DC (1981) Brainstorms: philosophical essays on mind and psychology. The MIT Press, Cambridge

Farrell S, Lewandowsky S (2010) Computational models as aids to better reasoning in psychology. Curr Dir Psychol Sci 19:329–335

Festinger L (1957) A theory of cognitive dissonance. Row, Peterson, Evanston

Fiedler K (2004) Tools, toys, truisms, and theories: some thoughts on the creative cycle of theory formation. Personal Soc Psychol Rev 8:123–131

Fiske ST (2004) Mind the gap: in praise of informal sources of formal theory. Personal Soc Psychol Rev 8:132–137

Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 103:650–669

Gigerenzer G, Todd PM (1999) Simple heuristics that make us smart. University Press, Oxford

Glöckner A, Betsch T (2011) The empirical content of theories in judgment and decision making: shortcomings and remedies. Judgm Decis Mak 6:711–721

Glöckner A, Hilbig BE, Jekel M (2014) What is adaptive about adaptive decision making? A parallel constraint satisfaction account. Cognition 133:641–666

Higgins ET (2004) Making a theory useful: lessons handed down. Personal Soc Psychol Rev 8:138–145

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47:263–292

Kirsch A (2019) A unifying computational model of decision making. Cognit Process. https://doi.org/10.1007/s10339-019-00904-3

Klein S (2014) What can recent replication failures tell us about the theoretical commitments of psychology? Theory Psychol 24:326–338

Krajbich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci 13:1292–1300

Marewski JN, Olsson H (2009) Beyond the null ritual: formal modeling of psychological processes. J Psychol 217:49–60

Myung IJ, Pitt MA (1997) Applying Occam's razor in modeling cognition: a Bayesian approach. Psychon Bull Rev 4:79–95

Myung IJ, Navarro DJ, Pitt MA (2006) Model selection by normalized maximum likelihood. J Math Psychol 50:167–179

Pitt MA, Myung IJ (2000) When a good fit can be bad. Trends Cognit Sci 6:421–425

Pitt MA, Myung IJ, Zhang S (2002) Towards a method for selecting among computational models for cognition. Psychol Rev 109:472–491

Platt JR (1964) Strong inference. Science 146:347–353

Popper KR (1934/2005) Logik der Forschung, 11th edn. Mohr Siebeck, Tübingen

Rieskamp J, Otto PE (2006) SSL: a theory of how people learn to select strategies. J Exp Psychol General 135:207–236

Roberts S, Pashler H (2000) How persuasive is a good fit? A comment on theory testing. Psychol Rev 107:358–367

Ross D (2019) Empiricism, sciences, and engineering: cognitive science as a zone of integration. Cogn Process. https://doi.org/10.1007/s10339-019-00916-z

Shultz TR, Lepper MR (1996) Cognitive dissonance reduction as constraint satisfaction. Psychol Rev 103:219–240

Thagard P (2012) The cognitive science of science: explanation, discovery, and conceptual change. The MIT Press, Cambridge

Van Lange PA, Kruglanksi AW, Higgins ET (2012) Theories of social psychology: an introduction. In: Van Lange PA, Kruglanksi AW, Higgins ET (eds) Handbook of theories of social psychology. Sage, London, pp 1–8

Vanpaemel W, Lee MD (2012) Using priors to formalize theory: optimal attention and the generalized context model. Psychon Bull Rev 19:1047–1056

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.