

# Editorial: Methodology in judgment and decision making research

Andreas Glöckner\*

Benjamin E. Hilbig<sup>†</sup>

## Abstract

In this introduction to the special issue on methodology, we provide background on its original motivation and a systematic overview of the contributions. The latter are discussed with correspondence to the phase of the scientific process they (most strongly) refer to: Theory construction, design, data analysis, and cumulative development of scientific knowledge. Several contributions propose novel measurement techniques and paradigms that will allow for new insights and can thus avail researchers in JDM and beyond. Another set of contributions centers around how models can best be tested and/or compared. Especially when viewed in combination, the papers on this topic spell out vital necessities for model comparisons and provide approaches that solve noteworthy problems prior work has been faced with.

Keywords: methodology, judgment and decision making.

## 1 Introduction

Methodology is one of the vital pillars of all science. Indeed, the question of *how* we go about our scientific quests—rather than *what* exactly we are investigating—has stimulated numerous debates and controversies over the past centuries. Mostly, this debate has served the common purpose of establishing certain standards which serve as a road map for scientists. Though disciplines and subfields vary greatly in their specific methodological standards, all share some degree of concern for such matters.

The field of psychology certainly is no exception. On the contrary, “[o]ne of the hallmarks of modern academic psychology is its methodological sophistication” (Rozin, 2009, p. 436). Methodological issues play a prominent role in the ongoing exchange and a growing number of contributions have recently addressed potential methodological problems inherent in the behavioral sciences (e.g., see the recent special issues in *Perspectives on Psychological Science* by De Houwer, Fiedler, & Moors, 2011; and Kruschke, 2011). Doubts have been raised concerning the subjects on which findings are typically based (Henrich, Heine, & Norenzayan, 2010), the approaches taken in theory development and testing (Gigerenzer, 1998; Henderson, 1991; Trafimow, 2003, 2009; Wallach & Wallach, 1994), the nature of the behavior assessed (Baumeister, Vohs, & Funder, 2007), or spe-

cific practices of data collection and questionable standards in data analysis (Dienes, 2011; Simmons, Nelson, & Simonsohn, in press; Wagenmakers, 2007; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011; Wetzels, et al., 2011), to name but a few examples.<sup>1</sup>

Also, there are several projects in development that attempt to coordinate collective action for solving fundamental methodological problems. The *Filedrawer Project* ([www.psychfiledrawer.org](http://www.psychfiledrawer.org)) provides an online archive of replication attempts to address the problem that “[m]ost journals [...] are rarely willing to publish even carefully conducted non-replications that question the validity of findings that they have published” ([www.psychfiledrawer.org/about.php](http://www.psychfiledrawer.org/about.php)); this problem, in turn can lead to publication biases (see Renkewitz, Fuchs, & Fiedler, 2011). In a similar vein, the *Reproducibility Project* (<http://openscienceframework.org>) aims to estimate the reproducibility of findings published in top psychological journals by conducting a collective, distributed attempt to replicate findings from a large sample of recently published papers. Needless to say, still other examples of papers and projects highlight methodological challenges and provide potential solutions. In a nutshell, all essentially hint at the continuous struggle for increasingly conclusive, robust, and general knowledge.

In our view, this struggle also goes on in the field of Judgment and Decision Making (JDM) research. Often enough, important advances in this area are motivated by methodological criticism. For example, of the many reactions to the recently proposed priority heuristic for risky choice (Brandstätter, Gigerenzer, & Hertwig, 2006) a substantial number raise methodological

We thank all authors for their stimulating contributions, the many reviewers who provided timely and vital feedback, and the editor-in-chief, Jon Baron, for making this special issue possible.

\*Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn (Germany). E-mail: [gloeckner@coll.mpg.de](mailto:gloeckner@coll.mpg.de).

<sup>†</sup>School of Social Sciences, University of Mannheim, 68131 Mannheim (Germany). E-mail: [hilbig@psychologie.uni-mannheim.de](mailto:hilbig@psychologie.uni-mannheim.de).

<sup>1</sup>For methodological debates in medicine which apply to psychological research see also Ioannidis (2005) and Ioannidis, Tatsioni, and Karassa (2010).

concerns pertaining to research strategy in general, the diagnosticity of tasks used, or the data analyses applied (e.g. Andersen, Harrison, Lau, & Rutström, 2010; Birnbaum, 2008a, 2008b; Birnbaum & LaCroix, 2008; Fiedler, 2010; Glöckner & Betsch, 2008; Hilbig, 2008; Regenwetter, Dana, & Davis-Stober, 2011; Regenwetter, Ho, & Tsetlin, 2007). Other theoretical controversies have similarly stimulated debate that largely centers around methodological issues (e.g., Brighton & Gigerenzer, 2011; Camilleri & Newell, 2011; Hilbig, 2010; Hilbig & Richter, 2011; Marewski, Schooler, & Gigerenzer, 2010; Pachur, 2011).

These examples and others demonstrate a need for an explicit and focused exchange of methodological arguments in JDM and potentially some room for improving common practices in this field. This assertion provided the main motivation for setting up a call for papers on methodology in JDM research. Aiming to keep our own agendas out of the early stages of development, we kept the initial call for papers deliberately broad. The gratifying upshot was an unexpectedly large number of interesting and important submissions.

Despite the breadth of the initial call, however, an early observation was that relatively few (if any) contributions dealt with issues in the philosophy of science or concerned methodological issues of theory formation and revision. Instead, the vast majority of manuscripts addressed issues of design and data analysis. This unequal distribution will become obvious in what follows: In this introduction to the special issue, we briefly discuss all contributions ordered by the stages of scientific discovery to which they (mostly) refer (see Figure 1).

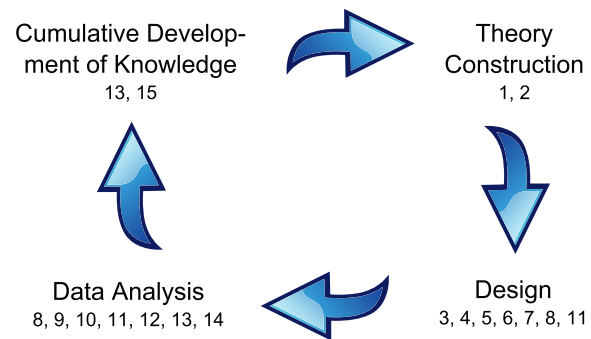
## 2 Overview of papers

In this overview, we commence with issues of theory construction, before we then turn to experimental design and measurement. Next, we discuss the papers pertaining to those steps that follow data collection, namely data analysis, and cumulative development of knowledge. Note, however, that several of the papers speak to more than one of these matters. As such, ordering and grouping the contributions in the current way should not be taken to imply that each paper relates to only one of the phases of scientific progress.

### 2.1 Theory construction

Two papers in this special issue discuss theory construction and theory development in the field of JDM (Glöckner & Betsch, 2011; Katsikopoulos & Lan, 2011). Following Poppers approach of critical rationalism, Glöck-

Figure 1: Overview of contributions.



- 1 Glöckner & Betsch: *Empirical Content of Theories*.
- 2 Katsikopoulos & Lan: *Herbert Simons Spell on JDM*.
- 3 Schulte-Mecklenbeck, et al.: *Process Tracing Methods*.
- 4 Franco-Watkins & Johnson: *Eye-moving window paradigm*.
- 5 Koop & Johnson: *Response dynamics measure*.
- 6 Doyle et al.: *Task selection for testing temporal discounting models*.
- 7 Murphy, et al.: *Social Value slider*.
- 8 Jekel, et al.: *Euclidian diagnostic task selection*.
- 9 Davis-Stober & Brown: *Normalized Maximum Likelihood*.
- 10 Moshagen & Hilbig: *Global fit test for model comparisons*.
- 11 Broomell, et al.: *Pairwise model comparisons*.
- 12 Lee & Newell: *Hierarchical Bayesian models*.
- 13 Matthews: *Bayesian hypothesis testing*.
- 14 Budescu & Johnson: *Mixed-model analysis of calibration*.
- 15 Renkewitz, et al.: *Publication biases in JDM*.

ner and Betsch advocate that scientific progress crucially necessitates that theories be formulated so as to comprise high empirical content, while being falsifiable. The authors point out some common drawbacks in corresponding theory formulation in JDM—especially a tendency towards formulation of weak theories. Also, for certain classes of JDM models, some remedies are suggested. More generally, observable shortcomings are partially attributed to a social dilemma structure (i.e., strictly maximizing personal interests would harm the collective interest to achieve scientific progress). It is suggested that the scientific community should agree upon a change in publication policies to overcome this dilemma structure.

Katsikopoulos and Lan take a historical perspective and discuss general developments in the field of JDM by investigating Herbert Simon's influence on current work. In a review of recent articles in the field, the authors demonstrate the strong influence that Simon's ideas had on today's thinking in JDM. Katsikopoulos and Lan also critically assess the way in which these ideas are treated in current work. In particular, the authors argue that integrative approaches for research on descriptive and prescriptive models are sought too seldom.

## 2.2 Design

Many of the contributions in this special issue focus on the steps between theory construction and data collection. That is, they concern the design stage, including the use of measurement methods, as well as the selection of appropriate tasks and stimuli.

### 2.2.1 Measurement methods

Schulte-Mecklenbeck, Kühberger, and Ranyard (2011) discuss classic and more recently developed process tracing methods and present examples of how these techniques can strongly aid development and testing of JDM process models. In a similar vein, Franco-Watkins and Johnson (2011) suggest applying an eye-moving window technique (i.e., information board in which information is revealed once it is looked at). They argue that this information board variant allows for combining the advantages of classic Mouselab techniques and eye-tracking; specifically, this method should allow for fast and effortless information acquisition, while ensuring that the researcher gains full insight on which information was looked up, for how long, and when.

A third paper proposing a new method to gain insight on cognitive processes was contributed by Koop and Johnson (2011). They suggest applying a measure of response dynamics which is based on analyzing different aspects of mouse-trajectories between a starting position and the option chosen. The underlying idea is that the attraction exerted by the non-chosen option will manifest itself in these trajectories (e.g., Spivey & Dale, 2006) and thus provides insight concerning the on-line formation of preferences. Overall, these different contributions commonly signify that the application and combination of classic and new methods will provide important insights concerning processes underlying judgment and decision making.

### 2.2.2 Diagnostic task selection

Another issue of research design discussed in several papers is the selection of tasks that allow for actually discriminating between theories or hypotheses. Doyle, Chen and Savani (2011) provide a method (using Excel-Solver) for selecting tasks that differentiate optimally between theoretical models of temporal discounting. They show how to construct tasks that make the rate parameters of prominent theories orthogonal or even inversely related.

In a rather different domain, Murphy, Ackermann and Handgraaf (2011) provide a method to measure social value orientation (Van Lange, 1999) by using a few highly diagnostic tasks in which participants distribute money between themselves and others. The innovative method is based on a slider format which—combined

with diagnostic tasks—makes data collection very efficient. Indeed, both approaches by Doyle et al. and Murphy et al. also seem promising in that they can probably be extended to other concepts relevant in JDM such as loss aversion, risk aversion etc.

Another contribution addresses the issue of diagnostic task selection from a somewhat different angle. Jekel, Fiedler and Glöckner (2011) provide a standard method for diagnostic task selection in probabilistic inference tasks. The suggested *Euclidian Diagnostic Task Selection method* increases the efficiency in research design and reduces the degree of subjectivity in task selection. Jekel et al. also provide a ready-made tool programmed in R that makes it easy to use the method in future research (see also Jekel et al., 2010). Overall, there is agreement that diagnostic task selection is crucial for model comparison and model testing.

## 2.3 Data analysis

The majority of papers in the special issue are concerned with core issues of data analysis, including contributions suggesting improved methods for model comparisons, demonstrating the advantages of Bayesian methods, or pointing to the advantages of mixed-model approaches.

### 2.3.1 Model comparisons

Several papers focus on methods for model comparisons. Davis-Stober and Brown (2011) describe how to apply a normalized maximum likelihood (NML) approach to strategy classification in probabilistic inference and risky choice. One of the crucial advantages is that NML takes into account models' overall flexibility instead of correcting for the number of free parameters only. The paper also illustrates how to test models assuming that decision makers do not stick to single strategies, but rather use a mixture of these.

Moshagen and Hilbig (2011) connect to the ideas discussed in Glöckner and Betsch, though focusing more on the importance of falsification. They show that comparing the fit of competing models can easily lead to entirely false conclusions once the true data-generating model is not actually among those considered (Bröder & Schiffer, 2003). As a remedy, they suggest including a test of absolute model fit which provides a chance for refuting false models.

Broomell, Budescu and Por (2011) show that the problem of overlapping model predictions (see also the contribution by Jekel et al.) can lead to biased conclusions in model comparisons and model competitions (see Erev, et al., 2010). The reason for this is that global measures of fit can hide the level of agreement between the predictions of various models. Broomell et al. propose the use

of more informative pair-wise model comparisons and demonstrate the advantages of such an approach. Also, the contribution by Jekel et al. discussed in the previous section adds insight on this matter by suggesting certain improvements in model comparisons. The same holds true for the hierarchical Bayesian approach put forward by Lee and Newell (2011) that is discussed in the next section.

### 2.3.2 Bayesian approaches

Another prominent issue concerns the application of Bayesian approaches and replacing classic methods of hypothesis testing by corresponding methods. Lee and Newell (2011) demonstrate the advantages of using hierarchical Bayesian methods for modeling search and stopping rules of decision strategies at the level of individuals. One of the core advantages over the strategy classification methods discussed above is that the hierarchical structure uses what has been learned about one subject for assisting inference for another one (“shrinkage”). Lee and Newell further show that their method will provide new insight on the nature of individual differences (e.g., in information search) which might also help to solve the debate between multi-strategy and uni-models for decision making (e.g., Newell, 2005).

In another paper on Bayesian methods, Matthews (2011) discusses potential advantages of replacing classic Fisherian and Neyman-Pearson hypothesis testing. He exemplifies that a reanalysis of previous studies when replacing classic t-tests by Bayesian t-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) can lead to strikingly different conclusions. The Bayesian approach allows for comparing mutually exclusive hypotheses on the same footage, thus avoiding the problems of p-values and allowing for evidence for the null hypothesis. In a long-term perspective, the Bayesian approach would also aid knowledge accumulation by considering the sum of previous research findings when setting the priors for later analyses. We hope that the paper inspires further constructive discussion concerning the clear advantages but also the remaining drawbacks of Bayesian statistics.

### 2.3.3 Mixed-model approaches

Budescu and Johnson (2011) suggest a model-based approach to improve the analysis of the calibration of probability judgments. In calibration research, judgments must be compared against event probabilities. However, event probabilities are often unknown. The authors show that aggregating over observations can lead to wrong conclusions and suggest using a model-based approach instead. Specifically, they put forward a mixed-model regression approach (simultaneously taking into account effects be-

tween and within subjects) to estimate event probabilities which are then compared against probability judgments to determine calibration. Similar to the hierarchical approach by Lee and Newell, one crucial advantage of this mixed-model based approach is that estimates for within- and between subjects effects are more stable because they profit from the larger underlying data basis.

## 2.4 Cumulative development of knowledge

There are two contributions in the special issue that—besides touching on questions of data analysis—also speak to the matter of cumulative development of knowledge. One is the above mentioned paper by Matthews (2011) on using Bayesian approaches. As mentioned above, replacing (or complementing) classic hypothesis testing by the Bayesian approach aids knowledge accumulation. In a second contribution, Renkewitz, Fuchs, and Fiedler (2011) address the important issue of publication biases. By exemplarily re-analyzing two JDM-specific meta-analyses, they demonstrate that publication biases are also present in JDM research. Such biases, in turn, will hinder appropriate cumulative development of knowledge. Indeed, severely distorted overall estimations of effect size—or even premature acceptance of the existence and stability of effects—can be the consequences. The authors discuss both specific methods to identify publication biases (in meta-analyses) and further provide recommendations on how changes in the overall standards and publication practices might counteract the problem identified.

## 3 Summary and conclusions

We are pleased to say that the 15 papers contained in this special issue avail many important insights in JDM methodology and provide helpful tools and suggestions which—in our view—will further improve the confidence we may have in our findings. Despite the fact that these 15 contributions are motivated by some methodological weaknesses in the field of JDM, it is also important to highlight that many of the problems tackled speak for the methodological sophistication of JDM research that is already in place. Of course, of those points raised in this special issue some are more and others less controversial. Indeed, our experience in handling these papers throughout the review process showed that some papers have more potential for debate than others. Nonetheless, the constructive way in which all contributions describe ways to overcome methodological weaknesses makes us optimistic that this issue might inspire further positive developments.

It seems as if the techniques and policies for improving our methodological standards are available. One of the foremost aims of the special issue was to inspire a more intense debate concerning these issues in order to improve the degree to which standards are shared within the community which is the basic requirement for their comprehensive enforcement. This is necessary for achieving scientific progress and overcoming social dilemma structures inherent in joint scientific discovery.

## References

(Those marked with \* are part of this special issue.)

- Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. (2010). Behavioral econometrics for psychologists. *Journal of Economic Psychology*, *31*, 553–576.
- 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*, 396–403.
- Birnbaum, M. H. (2008a). Evaluation of the priority heuristic as a descriptive model of risky decision making: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*, 253–260.
- Birnbaum, M. H. (2008b). New tests of cumulative prospect theory and the priority heuristic: Probability-outcome tradeoff with branch splitting. *Judgment and Decision Making*, *3*, 304–316.
- Birnbaum, M. H., & LaCroix, A. R. (2008). Dimension integration: Testing models without trade-offs. *Organizational Behavior and Human Decision Processes*, *105*, 122–133.
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). Making choices without trade-offs: The priority heuristic. *Psychological Review*, *113*, 409–432.
- Brighton, H., & Gigerenzer, G. (2011). Towards competitive instead of biased testing of heuristics: A reply to Hilbig and Richter (2011). *Topics in Cognitive Science*, *3*, 197–205.
- Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, *16*, 193–213.
- \* Broomell, S. B., Budescu, D. V., & Por, H.-H. (2011). Pair-wise comparisons of multiple models. *Judgment and Decision Making*, *6*, 821–831.
- \* Budescu, D. V., & Johnson, T. R. (2011). A model-based approach for the analysis of the calibration of probability judgments. *Judgment and Decision Making*, *6*, 857–869.
- Camilleri, A. R., & Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review*, *18*, 377–384.
- \* Davis-Stober, C. P., & Brown, N. (2011). A shift in strategy or “error”? Strategy classification over multiple stochastic specifications. *Judgment and Decision Making*, *6*, 800–813.
- De Houwer, J., Fiedler, K., & Moors, A. (2011). Strengths and limitations of theoretical explanations in psychology: Introduction to the special section. *Perspectives on Psychological Science*, *6*, 161–162.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274–290.
- \* Doyle, J. R., Chen, C. H., & Savani, K. (2011). New designs for research in delay discounting. *Judgment and Decision Making*, *6*, 759–770.
- Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., et al. (2010). A Choice Prediction Competition: Choices from Experience and from Description. *Journal of Behavioral Decision Making*, *23*, 15–47.
- Fiedler, K. (2010). How to study cognitive decision algorithms: The case of the priority heuristic. *Judgment and Decision Making*, *5*, 21–32.
- \* Franco-Watkins, A. M., & Johnson, J. G. (2011). Applying the decision moving window to risky choice: Comparison of eye-tracking and mousetracing methods. *Judgment and Decision Making*, *6*, 740–749.
- Gigerenzer, G. (1998). Surrogates for theories. *Theory & Psychology*, *8*, 195–204.
- Glöckner, A., & Betsch, T. (2008). Do people make decisions under risk based on ignorance? An empirical test of the priority heuristic against cumulative prospect theory. *Organizational Behavior and Human Decision Processes*, *107*, 75–95.
- \* Glöckner, A., & Betsch, T. (2011). The Empirical Content of theories in judgment and decision making: shortcomings and remedies. *Judgment and Decision Making*, *6*, 711–721.
- Henderson, D. K. (1991). On the testability of psychological generalizations (psychological testability). *Philosophy of Science*, *58*, 586–606.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.
- Hilbig, B. E. (2008). One-reason decision making in risky choice? A closer look at the priority heuristic. *Judgment and Decision Making*, *3*, 457–462.
- Hilbig, B. E. (2010). Reconsidering “evidence” for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, *17*, 923–930.
- Hilbig, B. E., & Richter, T. (2011). Homo heuristicus outnumbered: Comment on Gigerenzer and Brighton (2009). *Topics in Cognitive Science*, *3*, 187–196.

- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2, 697–701.
- Ioannidis, J. P., Tatsioni, A., & Karassa, F. B. (2010). Who is afraid of reviewers' comments? Or, why anything can be published and anything can be cited. *European Journal of Clinical Investigation*, 40, 285–287.
- \* Jekel, M., Fiedler, S., & Glöckner, A. (2011). Diagnostic task selection for strategy classification in judgment and decision making. *Judgment and Decision Making*, 6, 782–799.
- Jekel, M., Nicklisch, A., & Glöckner, A. (2010). Implementation of the Multiple-Measure Maximum Likelihood strategy classification method in R: Addendum to Glöckner (2009) and practical guide for application. *Judgment and Decision Making*, 5, 54–63.
- \* Katsikopoulos, K. V., & Lan, C.-H. (2011). Herbert Simon's spell on judgment and decision making. *Judgment and Decision Making*, 6, 722–732.
- \* Koop, G. J., & Johnson, J. G. (2011). Continuous process tracing and the Iowa Gambling Task: Extending response dynamics to multialternative choice. *Judgment and Decision Making*, 6, 750–758.
- Kruschke, J. K. (2011). Introduction to special section on bayesian data analysis. *Perspectives on Psychological Science* 6, 272–273.
- \* Lee, M. D., & Newell, B. R. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judgment and Decision Making*, 6, 832–842.
- Marewski, J. N., Schooler, L. J., & Gigerenzer, G. (2010). Five principles for studying people's use of heuristics. *Acta Psychologica Sinica*, 42, 72–87.
- \* Matthews, W. J. (2011). What would judgment and decision making research be like if we took a Bayesian approach to hypothesis testing? *Judgment and Decision Making*, 6, 843–856.
- \* Moshagen, M., & Hilbig, B. E. (2011). Methodological notes on model comparisons and strategy classification: A falsificationist proposition. *Judgment and Decision Making*, 6, 814–820.
- \* Murphy, R. O., Ackerman, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6, 771–781
- Newell, B. R. (2005). Re-Visions of rationality? *Trends in Cognitive Sciences*, 9, 11–15.
- Pachur, T. (2011). The limited value of precise tests of the recognition heuristic. *Judgment and Decision Making*, 6, 413–422.
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118, 42–56.
- Regenwetter, M., Ho, M.-H. R., & Tsetlin, I. (2007). Sophisticated approval voting, ignorance priors, and plurality heuristics: A behavioral social choice analysis in a Thurstonian framework. *Psychological Review*, 114, 994–1014.
- \* Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication biases in JDM research? *Judgment and Decision Making*, 6, 870–881.
- Rouder, J. N., Speckman, P. L., Sun, D. C., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Rozin, P. (2009). What kind of empirical research should we publish, fund, and reward? A different perspective. *Perspectives on Psychological Science*, 4, 435–439.
- \* Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, 6, 733–739.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (in press). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*.
- Spivey, M. J., & Dale, R. (2006). Continuous Dynamics in Real-Time Cognition. *Current Directions in Psychological Science*, 15, 207–211.
- Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review*, 110, 526–535.
- Trafimow, D. (2009). The theory of reasoned action: A case study of falsification in psychology. *Theory & Psychology*, 19, 501–518.
- Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77, 337–349.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of Psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100, 426–432.
- Wallach, L., & Wallach, M. A. (1994). Gergen versus the mainstream: Are hypotheses in social psychology subject to empirical test? *Journal of Personality and Social Psychology*, 67, 233–242.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t Tests. *Perspectives on Psychological Science*, 6, 291–298.