



Bayesian Revision vs. Information Distortion

J. Edward Russo^{1,2*}

¹ SC Johnson College of Business, Cornell University, Ithaca, NY, United States, ² Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

David R. Mandel,
Defence Research and Development
Canada, Canada

Reviewed by:

Norman Fenton,
Queen Mary University of London,
United Kingdom
Jean Baratgin,
Université Paris 8, France

*Correspondence:

J. Edward Russo
jer9@cornell.edu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 26 June 2018

Accepted: 06 August 2018

Published: 28 August 2018

Citation:

Russo JE (2018) Bayesian Revision
vs. Information Distortion.
Front. Psychol. 9:1550.
doi: 10.3389/fpsyg.2018.01550

The rational status of the Bayesian calculus for revising likelihoods is compromised by the common but still unfamiliar phenomenon of information distortion. This bias is the distortion in the evaluation of a new datum toward favoring the currently preferred option in a decision or judgment. While the Bayesian calculus requires the independent combination of the prior probability and a new datum, information distortion invalidates such independence (because the prior influences the datum). Although widespread, information distortion has not generally been recognized. First, individuals are not aware when they themselves commit this bias. In addition, it is often hidden in more obvious suboptimal phenomena. Finally, the Bayesian calculus is usually explained only with undistortable data like colored balls drawn randomly. Partly because information distortion is unrecognized by the individuals exhibiting it, no way has been devised for eliminating it. Partial reduction is possible in some situations such as presenting all data simultaneously rather than sequentially with revision after each datum. The potential dangers of information distortion are illustrated for three professional revision tasks: forecasting, predicting consumer choices from internet data, and statistical inference from experimental results. The optimality of the Bayesian calculus competes with people's natural desire that their belief systems remain coherent in the face of new data. Information distortion provides this coherence by biasing those data toward greater agreement with the currently preferred position—but at the cost of Bayesian optimality.

Keywords: Bayesian calculus, connectionism, desirability bias, forecasting, information distortion, likelihood updating, rationality, statistical inference

The information needed for nearly all important decisions falls into two categories, values and likelihoods. These decisions typically share two characteristics. First, both values and likelihoods are mainly subjective. Even objective information often requires a subjective evaluation of its decision impact. Second, important decisions usually involve the search for additional information that then drives the revision of the values and likelihoods. While there is no optimizing guidance for revising values, there is for likelihoods. That guidance is the Bayesian calculus¹.

¹The optimality of the Bayesian calculus presumes a fixed sample space of possible outcomes and their probabilities. In many real situations, this assumption is violated as, over time, additional outcomes become recognized or the probabilities of the original outcomes are altered (Baratgin and Politzer, 2006, 2010). The present claim of Bayesian optimality excludes such changes in the probability space. Respecting the difference between the effect of information to change the probabilities in a stable space and those that change the space itself, we use Bayesian revision for the former and reserve the term updating for the latter.

Bayesian revision combines a prior probability and the diagnostic value of a new datum (i.e., a unit of new information). The calculus of that combination requires that the prior probability and the datum contribute independently to the revised posterior probability. Unfortunately, a common phenomenon of likelihood revision can lead to the violation of that independence assumption.

This phenomenon is the predecisional distortion of information (Russo et al., 1996). It is illustrated by a study of whether to invest in a resort hotel (Russo and Yong, 2011). The sole investment criterion was risk as indexed by the probability of financial failure. Thus, the investment decision depended solely on the likelihood of failure. As information about the hotel was presented, the experimental participants revised the probability of the hotel's financial failure. That is, they repeatedly calculated a revised posterior probability after each new unit of information/datum.

The decision process, or equivalently the revision of posterior probabilities, was tracked by requiring two responses for each datum. The first was the judged diagnosticity of that datum. The second was the posterior probability, that is, the datum-driven revision of the likelihood of investing. The predecisional distortion of information is a bias in the evaluation of the new datum/information toward supporting whichever of the decision options is currently "in the lead." Consider potential investors who are leaning toward investing based on all the data seen so far (and captured by the prior probability). Then information distortion occurs when these investors bias their evaluation of the next datum toward investing. Conversely, if the same investors had been leaning toward not investing, information distortion (ID) is manifest as a biased interpretation of the next datum toward not investing. ID is a specific, process-explicated example of the many phenomena exhibiting a confirmation bias.

The impact of the predecisional distortion of information on the Bayesian calculus is depicted in **Figure 1**. The left panel illustrates the independent contributions of the prior and the datum to the posterior probability. The right panel adds the biasing influence of the prior on the datum. This is the influence of the current/prior leaning toward one option on the evaluation of the next datum/information. In their study of the resort investment decision, Russo and Yong (2011) reported significant information distortion (ID). This finding accords with similar risky decisions studied by DeKay et al. (DeKay et al., 2009, 2011; Glöckner and Herbold, 2011; Miller et al., 2013).

The phenomenon of information distortion (ID) during the revision of probabilities raises several questions. First, is ID widespread enough to affect a substantial number of revision tasks? Second, why has this bias not been recognized (and its consequences for the validity of the Bayesian calculus not been appreciated)? Third, what can be done to eliminate ID? The remainder of this article addresses these three questions. It concludes, first, with a consideration of where ID might undermine applications of the Bayesian calculus and, second, with a comment on the frequent clash between the ideal of normative criteria and the reality of human cognition (e.g., Thaler, 1992) of which the Bayesian calculus vs. ID is only one example.

HOW WIDESPREAD IS INFORMATION DISTORTION?

Reviews by DeKay (2015) and by Russo (2015) report the near universal presence of ID in decisions where the relevant information is acquired over time. Besides studies with college students and MTurk workers, ID has been found in decisions made by auditors, entrepreneurs, physicians, prospective jurors, and sales representatives.

In addition, ID is a systematic function of the prior commitment to the tentatively preferred course of action, as indexed by the prior probability. If that commitment is increased, ID rises in parallel (Polman and Russo, 2012).

ID is also persistent. Occasionally the new information/datum is so anti-leader that the posterior probability reflects a reversal of the leading option. In the above example, this might mean switching the tentative preference from investing to not investing. When such a preference reversal occurs, ID biases the evaluation of new information toward the new leading option, such as toward not investing. (However, see, Carlson et al., 2013, for residual traces of an initial preference).

WHY HAS INFORMATION DISTORTION NOT BEEN RECOGNIZED?

The presence of ID has gone unrecognized for multiple reasons. First, decision makers themselves are unaware that they distort new information. When ID is described to experimental subjects, their estimates of it in their own just-completed decision correlates essentially zero with their actual level of ID (Russo, 2015).

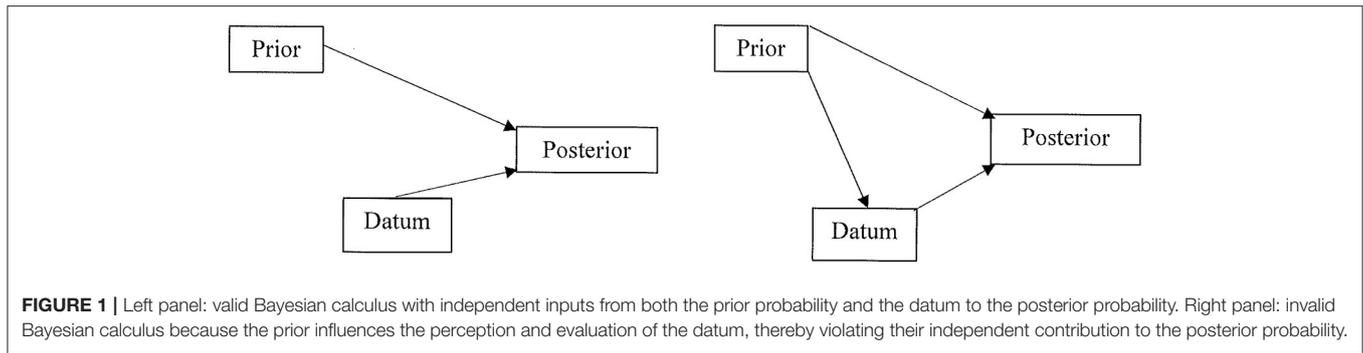
Second, sometimes ID is hidden among other biases. For example, ID is one among many possible causes of the desirability bias in which people overestimate the likelihood of a desired event (Russo and Corbin, 2016).

A third reason for the failure to detect the presence of ID may be peculiar to the canonical Bayesian setting. The familiar demonstrations of the Bayesian calculus have tended to use undistortable data, like contrastingly colored balls drawn randomly from an urn. It is impossible to distort the draw of 3 blue and 7 green balls, however strong may be the prior for one color.

In summary, the difficulty of recognizing ID in likelihood revision seems to have multiple causes. Some causes of ID seem omnipresent, such as the absence of self-awareness. Others operate only in certain situations, such as when conflated with other biases like desirability.

IS THERE REMEDIATION?

Remediation can be achieved, but only partially and in select circumstances. Consider first what may be the most obvious tactic for complete elimination, paying people to be accurate and, therefore, unbiased. (Meloy et al., 2006 see also Engel and Glöckner, 2013) found that incentives increased rather than decreased ID. Further, the negative impact of money held when



people were paid both for the accuracy of the decision and, more importantly, for the accuracy of the evaluation of each datum. (This anomalous result was caused by the positive mood induced by the incentive, Meloy, 2000).

A second potential path to remediation is to identify the cause of ID, which might suggest a method for its elimination. Russo et al. (2008) showed that ID is caused by the general desire for cognitive consistency (Gawronski and Strack, 2012). Specifically, decision makers want the datum (new information) to be consistent with the prior (which reflects all past information). In order to achieve greater consistency they distort the datum in the direction favored by the prior. Unfortunately, knowing that the goal of cognitive consistency drives ID does not reveal a method for ameliorating this bias besides the difficult task of increasing the tolerance for inconsistency. While methods for activating the consistency goal exist (Chaxel and Russo, 2015), those for deactivating it have proved elusive.

In spite of the failure of the above two common paths to amelioration, there have been some partial successes in reducing ID. The first relies on the simultaneous rather than sequential presentation of the data (Carlson et al., 2006). Simultaneous presentation reduces to near zero the delays between data that enable, even encourage, likelihood revision. Of course, such massed presentation does not prohibit spaced revision of the prior. Decision makers can still pause to consider the impact of each new datum before moving to the next information. Nonetheless, the simultaneous presentation of all data seems to inhibit such revision. One reason may be that the time and effort to process all the data at once are likely less than the cumulative total time and effort of several individual revisions.

A second tactic is creating precommitment to the diagnostic value of a datum. That is, each datum is evaluated prior to and, importantly, independent of a particular decision. Carlson and Pearo (2004) showed that if decision makers have knowledge of a datum outside the context of a decision, then ID is reduced almost to zero when that same datum appears during a decision.

Third, decision-making groups exhibit no ID so long as different members maintain opposing positions (Boyle et al., 2012). As long as some members favor one option while other members lean toward another option, there is sufficient debate on the pros and cons of each to suppress ID. This said, most decisions are not made in groups. Further and more worrying, once all members begins to lean toward the same option, ID grows to a level substantially above that of individuals.

In summary, ameliorative tactics are at least partially successful under some circumstances. However, no general strategy for eliminating ID has yet been devised.

THE RISK OF INFORMATION DISTORTION IN APPLICATIONS OF BAYESIAN INFERENCE

An appreciation of the value of Bayesian inference is increasing, as have the number and breadth of its applications. However, with this use of the Bayesian calculus comes the potential risk of contamination by ID. In some environments, such as those with undistortable data, ID can never taint Bayesian inference. However, in other likelihood revision tasks, a recognition of the possible presence of ID may improve the accuracy of Bayesian inference or at least prevent its misapplication.

The increased use of Bayesian inference/methods prompts a consideration of where ID might infect such applications. Three such areas are considered, with no claim to completeness or even representativeness. These are: the forecasting by experts of unique, complex events; the prediction of consumption behaviors from past consumption-related data; and Bayesian approaches to statistical inference. In all three cases, likelihood estimates based on new data/information are essential.

Forecasting by Experts

Although Bayesian methods for likelihood revision have generally not been used where only human judgments can provide a numerical evaluation of a datum, their use is increasingly likely. Consider forecasting, a professional task that has achieved recent success with the identification of “superforecasters” (Mellers et al., 2014, 2015). In the forecasting task, a datum is nearly always a unit of complex information. For instance, if the forecasted event is the reelection of Donald Trump in 2020, a positive datum might be the negotiated end to the Korean conflict of 1950-53. In contrast, a negative datum might be the criminal conviction of one of his inner circle. To apply Bayesian inference, forecasters would have to provide not only an explicit prior probability, as they often do now, but also a numerical judgment of the impact of each new datum, something not routinely required. The Bayesian calculus would then yield the revised posterior probability. In such forecasting, the risk of ID would emerge when experts’ commitment to a preferred

event, such as for or against Trump's re-election, biased their evaluation of a new datum (Mandel, 2008).

One familiar forecasting challenge is estimating the likelihood of the success of a new technology². However, what if the experts who must estimate the technology's success are also biased by ID? Consider the example of drug discovery, where pharmaceutical executives must decide whether to pursue the very expensive process of drug development and governmental approval. One of their challenges is that the only credible source of the likelihood of success is the expert scientists who developed the drug. Because they are often committed to its success, they may bias upward their estimated likelihood of eventual success. Yet to whom else can the decision-making executives turn for an informed likelihood of that success?

As typically practiced now, a forecaster need not provide an explicit likelihood ratio for a datum. If that were required, would it reveal a measurable effect of the prior on the evaluation of the datum in the form of ID? That is, might persistent ID in the human experts undermine the superior accuracy of the Bayesian calculus?

Prediction of Consumption

Bayesian models have a substantial history in commerce (e.g., Erdem and Keane, 1996). More specifically, their use by market researchers relies on both past consumption and internet product search to predict future consumption and, more recently, further search (e.g., Ching et al., 2013; Fong, 2017). However, even as these "consumer learning" models become more sophisticated, they tend to be revised between purchases only by the knowledge of what consumer has searched. The evaluation of that acquired information, along with the possible presence of ID, is not included in the models. How much might the predictive accuracy of such models be improved if they accounted for the biasing influence of ID?

Statistical Inference/Analysis

A third and growing domain/area of application of the methods of Bayesian inference is statistical tests of scientific hypotheses. See, for example, the set of papers introduced by Vandekerckhove et al. (2018). As has become well established, using the results from current data to determine when to stop collecting additional data (data-dependent optional stopping) risks invalidating the *p*-values and confidence intervals of classical hypothesis testing. This risk of invalidation has prompted the shift to pre-registering the plan of data collection. Advocates for Bayesian methods claim the elimination of such risks. This claim, in turn, requires the absence of bias in experimenter judgments during the process of inference from collected data.

An analysis of what researchers actually do suggests that this claim may be too strong. For instance, Dunbar (1995, 1999) observed the discussions of research biologists. He found that among the first potential explanations for anomalous data was error in the data collection method (instead of the invalidity of their proffered hypotheses). Surely the same reaction is plausible

in the experimental social sciences where data are frequently direct responses from human subjects. Thus, once experimenters who are committed to one hypothesis must judge the validity of their own data, ID may occur.

THE RATIONALITY OF BAYESIAN INFERENCE VS. THE MULTIPLE GOALS OF COGNITIVE PROCESSING

The Bayesian calculus belongs to the dominant class of decision theories that rely on the unbiased evaluation of information. Indeed, who would want such a theory if it accommodated rather than rejected a bias like ID? Nonetheless, such theories exist, albeit with descriptive rather than normative status. The most relevant may be connectionist models, which not only accept the ID bias but seem to need it. In these models, new information exerts a bidirectional influence on an existing network of related beliefs. A bidirectional process enables them to accommodate ID as the natural (to these models) influence of a current belief on the evaluation of new information (Holyoak and Simon, 1999; Glöckner and Herbold, 2011). This bidirectional influence contributes to the desired goal of a more coherent and stable system of beliefs as it accommodates to the new information.

Connectionist models do not claim rationality. Nonetheless, the goals of internal coherence and network stability are desirable outcomes of the processing of new information. Thus, an undesirable bias like ID becomes necessary to achieving the desirable ends of coherence and stability (Engel and Glöckner, 2013). Nonetheless, the prominence of connectionist models has tended to obscure the situations where coherence pays the price of tolerating biases like ID.

The descriptive value of connectionist theories coupled with the appeal of the goals that they achieve stands against the normative value of the Bayesian calculus. In tasks where that calculus is needed, such as forecasting, the admittedly desirable goals that drive connectionist dynamics must be sacrificed. Instead, techniques of cognitive engineering need to be developed to counter the natural associative mechanisms that yield ID and other phenomena that compromise the Bayesian calculus. Such help may be found in the techniques that enable superforecasting, such as structured methods for eliciting uncertainty estimates and for statistical reasoning. Two promising examples of structured methods for elicitation are the CHAMPS KNOW training that Mellers et al. (2014) used in the IARPA ACE tournament and Mandel (2015) training of intelligence analysts in Bayesian reasoning using natural sampling trees. Also valuable are disconfirmatory challenges from multiple individuals. When individuals work alone to predict an event of only personal relevance, the techniques of superforecasters may have limited application. Nonetheless, when a task is important enough, such as predicting the success of new technologies like drugs, a team may apply the multiple techniques of superforecasting to achieve Bayesian rationality.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

²By the definition of "new," there are no baserates to provide historical evidence of the technology's success. The absence of such baserates precludes the tactic of taking an "outside" view as advocated by Lovall and Kahneman (2003).

REFERENCES

- Baratgin, G., and Politzer, G. (2010). Updating: a psychologically basic situation of probability revision. *Think. Reason.* 16, 253–287. doi: 10.1080/13546783.2010.519564
- Baratgin, J., and Politzer, G. (2006). Is the mind Bayesian? The case for agnosticism. *Mind Soc.* 5, 1–38. doi: 10.1007/s11299-006-0007-1
- Boyle, P. J., Hanlon, D., and Russo, J. E. (2012). The value of task conflict to group decisions. *J. Behav. Decis. Mak.* 25, 217–227. doi: 10.1002/bdm.725
- Carlson, K. A., Meloy, M. G., and Miller, E. G. (2013). Goal reversion in consumer choice. *J. Consum. Res.* 39, 918–930. doi: 10.1086/666471
- Carlson, K. A., Meloy, M. G., and Russo, J. E. (2006). Leaderdriven primacy: using attribute order to affect consumer choice. *J. Consum. Res.* 32, 513–518. doi: 10.1086/500481
- Carlson, K. A., and Pearo, L. (2004). Limiting predecisional distortion by prior valuation of attribute components. *Organ. Behav. Hum. Decis. Proc.* 94, 48–59. doi: 10.1016/j.obhdp.2004.02.001
- Chaxel, S., and Russo, J. E. (2015). “Cognitive consistency: cognitive and motivational perspectives,” in *Neuroeconomics, Judgment, and Decision Making* eds A. W. Evan and F. R. Valerie (New York, NY: Psychology Press), 29–48.
- Ching, A. T., Erdem, T., and Keane, M. P. (2013). Learning models: an assessment of progress, challenges, and new developments. *Market. Sci.* 32, 913–938. doi: 10.1287/mksc.2013.0805
- DeKay, M. L. (2015). Predecisional information distortion and the self-fulfilling prophecy of early preferences in choice. *Curr. Dir. Psychol. Sci.* 24, 405–411. doi: 10.1177/0963721415587876
- DeKay, M. L., Patino-Echeverri, D., and Fischbeck, P. S. (2009). Distortion of probability and outcome information in risky decisions. *Organ. Behav. Hum. Decis. Proc.* 109, 79–92. doi: 10.1016/j.obhdp.2008.12.001
- DeKay, M. L., Stone, E. R., and Miller, S. A. (2011). Leader-driven distortion of probability and payoff information affects choices between risky prospects. *J. Behav. Decis. Mak.* 24, 394–411. doi: 10.1002/bdm.699
- Dunbar, K. (1995). “How scientists really reason: Scientific reasoning in real-world laboratories,” in *The Nature of Insight*, eds R. J. Sternberg and J. Davidson (Cambridge, MA: MIT Press), 365–395.
- Dunbar, K. (1999). “How scientists build models: *in vivo* science as a window on the scientific mind,” in *Model-based Reasoning in Scientific Discovery*, eds L. Magnani, N. Nersessian, and P. Thagard (Boston, MA: Plenum Press), 89–98.
- Engel, C., and Glöckner, A. (2013). Role-induced bias in court: an experimental analysis. *J. Behav. Decis. Mak.* 26, 272–284. doi: 10.1002/bdm.1761
- Erdem, T., and Keane, M. P. (1996). Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. *Market. Sci.* 15, 1–20. doi: 10.1287/mksc.15.1.1
- Fong, N. M. (2017). How targeting affects consumer search. *Manage. Sci.* 63, 2353–2364. doi: 10.1287/mnsc.2016.2447
- Gawronski, B., and Strack, F. (2012). *Cognitive Consistency: A Fundamental Principle in Social Cognition*. New York, NY: The Guilford Press.
- Glöckner, A., and Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: evidence for compensatory strategies based on automatic processing. *J. Behav. Decis. Mak.* 24, 71–98. doi: 10.1002/bdm.684
- Holyoak, K. J., and Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *J. Exp. Psychol. Gen.* 128, 3–31. doi: 10.1037/0096-3445.128.1.3
- Lovalló, D., and Kahneman, D. (2003). Delusions of success. *Harv. Bus. Rev.* 81, 56–63. doi: 10.1225/R0307D
- Mandel, D. R. (2008). Violations of coherence in subjective probability: a representational and assessment process account. *Cognition* 106, 130–156. doi: 10.1016/j.cognition.2007.01.001
- Mandel, D. R. (2015). Instruction in information structuring improves Bayesian judgment in intelligence analysts. *Front. Psychol.* 6:387. doi: 10.3389/fpsyg.2015.00387
- Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., et al. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspect. Psychol. Sci.* 10, 267–281. doi: 10.1177/1745691615577794
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., et al. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychol. Sci.* 25, 1106–1115. doi: 10.1177/0956797614524255
- Meloy, M. G. (2000). Mood-driven distortion of product information. *J. Consum. Res.* 27, 345–359. doi: 10.1086/317589
- Meloy, M. G., Russo, J. E., and Miller, E. G. (2006). Monetary incentives and mood. *J. Market. Res.* 43, 267–275. doi: 10.1509/jmkr.43.2.267
- Miller, S. A., DeKay, M. L., Stone, E. R., and Sorenson, C. M. (2013). Assessing the sensitivity of information distortion to four potential influences in studies of risky choice. *Judgm. Decis. Mak.* 8, 662–677. Available online at: <https://econpapers.repec.org/RePEc:jdm:journl:v:8:y:2013:i:6:p:662-677>
- Polman, E., and Russo, J. E. (2012). Reconciling competing beliefs during decision making. *Organ. Behav. Hum. Decis. Proc.* 119, 78–88. doi: 10.1016/j.obhdp.2012.05.004
- Russo, J. E. (2015). “The Predecisional Distortion of Information,” in *Neuroeconomics, Judgment, and Decision Making*, eds A. W. Evan and F. R. Valerie (New York, NY: Psychology Press), 91–110.
- Russo, J. E., and Corbin, J. (2016). Not by desire alone: the role of cognitive consistency in the desirability bias. *Judg. Dec. Makin.* 11, 449–459. Available online at: <https://econpapers.repec.org/RePEc:jdm:journl:v:11:y:2016:i:5:p:449-459>
- Russo, J. E., Kurt, A. C., Margaret, G. M., and Kevyn Y. (2008). The goal of consistency as a cause of information distortion. *J. Exp. Psychol. Gen.* 137, 456–470. doi: 10.1037/a0012786
- Russo, J. E., Medvec, V. H., and Margaret, G. M. (1996). The distortion of information during decisions. *Organ. Behav. Hum. Decis. Proc.* 66, 102–110. doi: 10.1006/obhd.1996.0041
- Russo, J. E., and Yong, K. (2011). The distortion of information to support an emerging assessment of risk. *J. Economet.* 162, 132–139. doi: 10.1016/j.jeconom.2010.07.004
- Thaler, R. L. (1992). *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. New York, NY: Free Press.
- Vandekerckhove, J., Rouder, J. N., and Kruschke, J. K. (2018). Editorial: bayesian methods for advancing psychological science. *Psych. Bull. Rev.* 25, 1–4. doi: 10.3758/s13423-018-1443-8

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Russo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.