A Unified Model of Entropy and the Value of Information

Jonathan D Nelson (jonathan.d.nelson@gmail.com)[1,2]
Vincenzo Crupi (vincenzo.crupi@unito.it)[3]
Björn Meder (meder@mpib-berlin.mpg.de)[1]
Gustavo Cevolani (g.cevolani@gmail.com)[2]
Katya Tentori (katya.tentori@unitn.it)[4]

[1]Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany
[2]School of Psychological Sciences, University of Surrey, Guildford, UK
[3]Center for Logic, Language and Cognition, Department of Psychology and Education, University of Turin, Italy
[4]Center for Mind and Brain Sciences, University of Trento, Italy

Keywords: uncertainty; entropy; information; information gain, probability gain

Abstract

Notions of entropy and uncertainty are fundamental to many domains, ranging from the philosophy of science to physics. One important application is to quantify the expected usefulness of possible experiments (or questions or tests). Many different entropy models could be used; different models do not in general lead to the same conclusions about which tests (or experiments) are most valuable. It is often unclear whether this is due to different theoretical and practical goals or are merely due to historical accident. We introduce a unified two-parameter family of entropy models that incorporates a great deal of entropies as special cases. This family of models offers insight into heretofore perplexing psychological results, and generates predictions for future research.

Uncertainty and Information

Notions of entropy and uncertainty are fundamental to many domains, ranging from the philosophy of science to physics. One important application of uncertainty is to quantify the expected usefulness of possible experiments (or questions or tests). Lindley (1956) suggested that an experiment's usefulness could be quantified in terms of how much it reduces expected Shannon (1948) uncertainty about the possible states of the world. This idea has proven useful in psychological models (Oaksford & Chater, 1994) as well. In a psychological context, the possible states could be the different categories that an object might belong to, and "experiments" could be a child's queries to learn more about the category. Other entropy models, such as Quadratic entropy (Crupi & Tentori, 2014) or Bayes's error (Baron, Beattie & Hershey, 1998; Crupi, Tentori & Lombardi, 2009) could also be used. Different models do not in general lead to the same conclusions about which tests (or experiments) are most valuable (Nelson, 2005, 2008, 2009).

What kind of entropy model best characterizes people's goals in searching for information? Some data suggest that reduction in Bayes's error (probability gain) is a more plausible intuitive model than reduction in Shannon entropy (Nelson, McKenzie, Cottrell & Sejnowski, 2010; Meder & Nelson, 2012). Probability gain appears to have its own limitations, however, as it does not show a preference for questions with close to a 50:50 split in 20-questions games (Nelson, Divjak, Gudmundsdottir, Martignon & Meder, 2014).

Many different ideas of important axioms for entropy measures have been proposed (Csiszár, 2008). Interestingly, particular entropy measures have been predominant in particular research areas, and it is often unclear whether this is due to different theoretical and practical goals or are merely due to historical accident.

Is there any possibility for a formal model of uncertainty that would be able to describe people's behavior across a wide variety of tasks? Could such a model also have theoretically desirable properties?

Entropy is often thought of as expected surprise. But (1) what constitutes surprise, and (2) what constitutes an expectation? Depending on how surprise and expectation are defined, different entropy measures result. Combining these two ideas, we show that many entropy measures, including Hartley (1928), Shannon (1948) and Quadratic entropy, and the families of Tsallis (1988), Rényi (1961), and Arimoto (1971) entropies, can all be derived as special cases in the Sharma-Mittal (1975) framework for entropy measures.

Figure 1 depicts the Sharma-Mittal space of entropy measures graphically, where the horizontal axis (the order r) specifies the type of averaging function, and the vertical axis (the degree t) specifies the surprise function. A number of heuristic ideas of uncertainty, for instance the number of possibilities, and whether or not you know for sure (analogous to a Popperian formulation, Popper, 1959), also arise as special cases in this framework.

Can psychological insight be derived from this formalism? We show that many heretofore disparate-seeming empirical results and normative desiderata can be accommodated by specific entropy measures within this formalism. Importantly, this framework affords more than a post hoc story; novel predictions can be derived for future experiments, to better characterize the psychological bases of uncertainty and information.

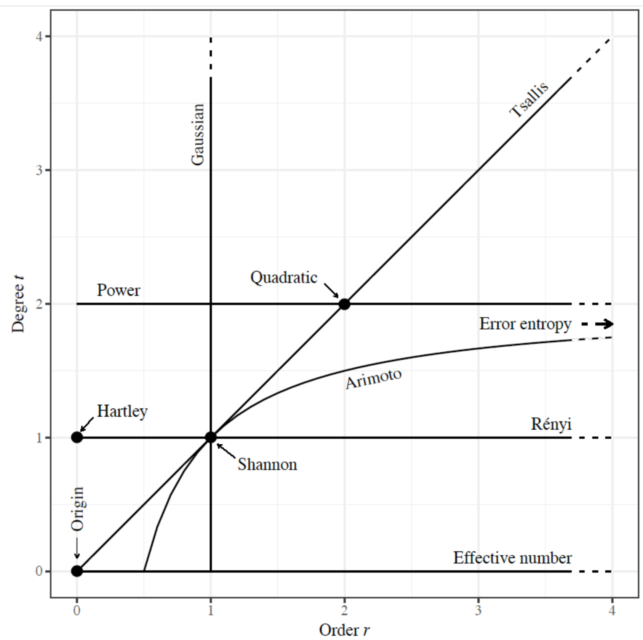

Figure 1: The Sharma-Mittal framework

Acknowledgments

This research was supported by grants CR 409/1-2, NE 1713/1-2, and ME 3717/2-2 from the Deutsche Forschungsgemeinschaft as part of the “New Frameworks of Rationality” priority program (SPP 1516). We thank Nick Chater, Laura Martignon, Andrea Passerini, and Paul Pedersen for helpful comments and discussions.

References

Arimoto S. (1971). Information-theoretical considerations on estimation problems. *Information and Control*, *19*, 181-194.

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, *42*, 88–110. doi: 10.1016/0749-5978(88)90021-0

Crupi, V., & Tentori, K. (2014). State of the field: Measuring information and confirmation. *Studies in History and Philosophy of Science*, *47*(*C*), 81–90. doi: 10.1016/j.shpsa.2014.05.002

Crupi, V., Tentori, K., & Lombardi, L. (2009). Pseudodiagnosticity revisited. *Psychological Review*, *116*, 971–985. doi: 10.1037/a0017050

Csiszár, I. (2008). Axiomatic characterizations of information measures. *Entropy*, *10*, 261–273. doi: 10.3390/e10030261

Hartley R. (1928). Transmission of information. *Bell System Technical Journal*, *7*, 535-563.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, *27*, 986–1005. doi: 10.1214/aoms/1177728069

Meder, B., & Nelson, J. D. (2012). Information search with situation-specific reward functions. *Judgment and Decision Making*, *7*, 119–148.

Nelson, J. D. (2005). Finding useful questions: on Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*, 979–999. doi: 10.1037/0033-295X.112.4.979

Nelson, J. D. (2008). Towards a rational theory of human information acquisition. In N. Chater and M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Rational Models of Cognition* (pp. 143–163). Oxford: Oxford University Press.

Nelson, J. D. (2009). Naïve optimality: Subjects’ heuristics can be better-motivated than experimenters’ optimal models. *Behavioral and Brain Sciences*, *32*, 94–95. doi: 10.1017/S0140525X09000405

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition*, *130*, 74–80. doi: 10.1016/j.cognition.2013.09.007

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science*, *21*, 960–969. doi: 10.1177/0956797610372637

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631. doi: 10.1037/0033-295X.101.4.608

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Rényi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 547–561.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423, 623–656.

Sharma, B. & Mittal, D. (1975). New non–additive measures of entropy for discrete probability distributions. *Journal of Mathematical Sciences (Delhi)*, *10*, 28-40.

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52, 479–487. doi: 10.1007/BF01016429