Quantifying Cognitive Diversity in Humans with Complex Causal Belief Systems: the Heterogeneity of Bayesian Prior Distributions.

Johannes Castner

August 27, 2013

1 Notes

1.1 undigested

From my discussion with John Miller: important first message is that PhD thesis are small! More particularly, I should find substantively interesting Bayesian networks and write a paper about a distance measure for such graphs (that could be my main paper, if all else fails). Next, I should find a few pages from the Congressional Record, hand code them and then compare to the output from the causal-belief-catcher. From discussion with Matt Jackson: the two dimensions of individual beliefs (uncertainty and direction of causal effect) are very different in that higher degrees of uncertainty could lead to more cooperation, as the agents are less dogmatic, whereas greater differences in directionality should lead to more polarization (this could be an interesting model and it could lead to two measures of diversity with different predictions). Thus, there should be two separate difference measures: one on difference in ambiguity and one in direction. Try with example networks. This may be related to a small model, where the same information leads to divergence of opinions, rather than convergence. From conversation with Rajiv Sethi: make assumptions very clear. One way to get the μ_+ and σ_+, for example, is to integrate across all possible distributions, such that a negative, or 0 belief is rejected. For each statement, there is a threshold, τ, such that

all other possibilities are rejected and thus the statement is supported and produced. All values of μ, given a particular σ, times their probability, to the right of that threshold (in the case of positive believes) must be integrated. From this video (`http://vimeo.com/66820855#`): Democracy as problem solving ...this seems to be the way I should conceptualize things? There is some bias coming from statements being public, rather than private. Bias toward outrageous stories, rather than the banal ...difficulty finding what is in people's heads. Also, homophily vs. social influence.

1.2   digestions

Griffith: Most human inferences are guided by background knowledge, and cognitive models should formalize this knowledge and show how it can be used for induction.

Me: Induction is important for decision making and thus, differences in models to be used for induction are differences in models for decision making. Thus, differences in inductive models (represented as Bayesian Networks) are the relevant cognitive differences political or economic theorists should be concerned with, together with differences in goals and values. If values are coded as causal relationships from the nodes that are endowed with "intrinsic value" (as opposed to instrumental value) to well-being (utility), then they can be included in this framework and a person changes her values when she determines, through new information or persuasion, that previously considered "intrinsically valuable" variables do not "cause" well-being.

Griffith: ...the prior distribution used by a Bayesian model is critical, since an appropriate prior can capture the background knowledge that humans bring to a given inductive problem.

Me: When making a model of someone else's mind (a politician, for example), then, everything that is known about that person's beliefs must be included in the model, as well as some additional assumptions (constraints on the prior distribution) about that person's belief system. It is a model of a model (a meta-model) which can be formulated entirely as a Bayesian Network. This can be done in such a way that both the theorist's beliefs about the person's mental model which is being under consideration, as well as the mental model itself can be simultaneously updated when new information about both, the world and the mental model become available; we can learn more about how the mind of a given person works, while this person learns about the world about which he forms his beliefs. In other words, as-

sumptions about a person's mind [1] can be included in the prior distribution (the joint distribution of a set of considered variables), which is a model of a person's causal theory of the world and can be specified as Hierarchical Bayesian model.

Differences in causal models can not be thought of entirely as differences in Bayesian priors which would (through updating) lead to an eventual convergence of beliefs in the face of enough evidence. Indeed, differences in belief systems, represented as Bayesian Networks, include heterogeneities in the likelihood functions of conditional distributions, where the structural models and not only the parameters are heterogeneous. To be precise, model selection, as opposed to parameter selection (estimation), has many components (i.e, the exact functional forms of all relations between variables and the type of stochasticity of the joint distribution of all variables), but the focus of the current work is on the dependencies between all relevant variables or what one might call the dependency structure of the causal model. Thus, rational individuals with differentially structured causal models of the world might come to diverge (rather than converge) on their opinions as the amount of data increases, regardless of the exact stochastic components or functional forms in their models. Politicians, for example, might come to increasingly disagree about their policy recommendations as evidence from the political world mounts, even when they agree on all normative considerations (i.e. when their "values" are the same) because they differ in terms of the structures of their belief systems.

2 Bayesian Models of Cognition

Quoted directly from Bayesian models of cognition (Thomas L. Griffiths, Charles Kemp and Joshua B. Tenenbaum xx): formulated as in Marrs (1982) level of "computational theory", rather than the algorithmic or process level that characterizes more traditional cognitive modelling paradigms, as described in other chapters of this volume: connectionist networks (see the chapter by McClelland), exemplar-based models (see the chapter by Logan),

---

[1] For example, there must be constraints on believed causal effects, such that when a person says that a causal effect from variable A to B is positive, in the absence of an explicit statement about the magnitude of such an effect, a prior on the magnitude of this believed effect must be specified (it can not be infinite for the obvious reason that all calculations would be meaningless in that case).

production systems and other cognitive architectures (see the chapter by Taatgen and Anderson), or dynamical systems (see the chapter by Shoener). Algorithmic or process accounts may be more satisfying in mechanistic terms, but they may also require assumptions about human processing mechanisms that are no longer needed when we assume that cognition is an approximately optimal response to the uncertainty and structure present in natural tasks and environments (Anderson, 1990). Finding effective computational models of human cognition then becomes a process of considering how best to characterize the computational problems that people face and the logic by which those computations can be carried out (Marr, 1982). The more complex methods can support multiple hierarchically organized layers of inference, structured representations of abstract knowledge, and approximate methods of evaluation that can be applied efficiently to data sets with many thousands of entities. For the first time, we now have practical methods for developing computational models of human cognition that are based on sound probabilistic principles and that can also capture something of the richness and complexity of everyday thinking, reasoning and learning.

Me: A Causal Bayes Net is used to represent a person's belief system, the tool with which to interpret events, and as such it is both the result and the cause of the previous structure of the net. Thus, given two distinct belief systems it could easily be the case that with the revelation of new data (where both agents are exposed to the same sensory input), the two belief systems diverge (the facts are increasingly interpreted differently) and a good measure of distance for any two belief systems should measure the degree to which they diverge/converge when new facts are released at some rate τ (which can be set to 1).

Griffith et al: Bayesian inference can also be applied in contexts where there are (uncountably) infinitely many hypotheses to evaluate–a situation that arises often.

Me: I want to apply this framework to the case where people's hypothesis are about the signs of causation in a causal system (priors and posteriors over structural models of the world).

Griffith: The posterior distribution over θ contains more information than a single point estimate: it indicates not just which values of θ are probable, but also how much uncertainty there is about those values. Collapsing this distribution down to a single number discards information, so Bayesians prefer to maintain distributions wherever possible (this attitude is similar to Marrs (1982, p. 106) "principle of least commitment").

4

Me: How to encode causal statements as distributions and how to take differences of two conditional distributions, then, become the questions of this work. Instead of maintaining information, I must find a way to take less information (coarse statements about the sign of causation) and turn them into statements about distributions (how much information is there in such coarse statements and how can I turn them into Bayesian priors?)

Griffith: Crucially for modelers interested in higher-level cognition, conjugate priors cannot capture knowledge that the causal process generating the observed data could take on one of several qualitatively different forms. A major area of current research in Bayesian statistics and machine learning focuses on building more complex models that maintain the benefits of working with conjugate priors, building on the techniques for model selection that we discuss next (e.g., Neal, 1992, 1998, Blei, Griffiths, Jordan, & Tenenbaum, 2004; Griffiths & Ghahramani, 2005). Hypotheses that differ in their complexity can be compared directly using Bayes' rule, once they are reduced to probability distributions over the observable data (see Kass & Raftery, 1995).

Me: does that mean that comparing two causal graphs (once converted to Bayes Nets) can be seen as two complex hypothesis that can be directly compared, using model selection techniques?

3   Causal Beliefs and Baysian Nets

Griffith: Graphical models provide an efficient and intuitive framework for working with high-dimensional probability distributions, which is applicable when these distributions can be viewed as the product of smaller components defined over local subsets of variables. A graphical model associates a probability distribution with a graph. The nodes of the graph represent the variables on which the distribution is defined, the edges between the nodes reflect their probabilistic dependencies, and a set of functions relating nodes and their neighbors in the graph are used to define a joint distribution over all of the variables based on those dependencies. If the edges indicate the direction of a dependency, the result is a directed graphical model. Our focus here will be on directed graphical models, which are also known as Bayesian networks or Bayes nets (Pearl, 1988). Bayesian networks can often be given a causal interpretation, where an edge between two nodes indicates that one node is a direct cause of the other, which makes them particularly appealing

for modeling higher-level cognition.

Griffith: A Bayesian Network represents the probabilistic dependencies relating a set of variables. If an edge exists from node A to node B, than A is referred to as a "parent" of B, and B is a "child" of A. This genealogical relation is often extended to identify the "ancestors" and "descendants" of a node. The directed graph used in a Bayesian network has one node for each random variable in the associated probability distribution and is constrained to be *acyclic*.

Me: constraining the directed belief graph to be acyclic is perhaps too demanding, as that seems to artificially preclude any kind of beliefs in feedbacks, such as are equilibria or runaway effects.

Griffith: The edges express the probabilistic dependencies between the variables in a fashion consistent with the *Markov Condition* (Pearl, 1988; Spirtes, Glymoure & Schienes, 1993). As a consequence of the Markov Condition, any Bayesian Network specifies a canonical factorization of a full joint probability distribution into the product of local conditional distributions, one for each variable conditioned on its parents.

Me: My obvious question here is whether and how the assumption that the dependency graph is acyclic can be dropped so that the above factorization can still be accomplished.

3.1   Causal Bayesian Nets

Griffith: In a standard Bayesian Network, edges between variables indicate only statistical dependencies between them. However, recent work has explored the consequences of augmenting directed graphical models with a stronger assumption about the relationship indicated by edges: that they indicate direct causal relationships (Pearl, 2000; Spirtes, Glymoure & Schienes, 1993). This assumption allows causal graphical models to represent not just the probabilities of events that one might observe, but also the probabilities of events that one can produce through intervening on a system.

Me: This makes causal Bayesian Networks very relevant as representations of belief systems of politicians, who by their very professions are required to hold beliefs over the probabilities of events that certain interventions (policies) can produce.

Griffith: The inferential implications of an event differ strongly, depending on whether it was observed passively or under conditions of intervention.

Me: For instance, one politician may believe that an increase in the number of college students who seek to borrow in order to finance their studies would cause an increase in the interest rates on student loans, but if other policy makers intervened by holding interest rates on student loans fixed, the unchanged interest rates would obviously not be surprising and would hardly contradict the politicians beliefs about the matter.

Griffith: In causal graphical models, the consequences of intervening on a particular variable can be assessed by removing all incoming edges to that variable and performing probabilistic inference in the resulting "mutilated" model (Pearl, 2000).

Me: In the above example, the incoming edge from the number of students seeking to borrow to the interest rate on student loans is removed by the intervention and thus, the intervention acts so as to break this causal relationship, rendering the two variables statistically independent. Thus, in the presence of the intervention, the current interest rate on student loans can not be used as a predictor of the current demand for student loans.

Griffith: Several recent papers have investigated whether people are sensitive to the consequences of intervention, generally finding that people differentiate between observational and interventional evidence appropriately (Hagmayer, Sloman, Lagnado & Waldmann, in press; Lagnado & Sloman, 2004; Steyvers et al., 2003).

4   Introduction

Many objects in nature have theoretical analogues which are, either in part or in whole, represented in the form of (possibly directed and/or weighted) Graphs. Among these objects are human brains, food webs (ecosystems), socio-economic and political structures in human societies, city roads and power-grids. Often, theories, or empirical tests of theories, will be concerned with the diversity of such objects (Weitzman 1992, Page XX), or with the partitioning of such objects into similarity clusters (Grimmer and King 2011). Both, the calculation of a diversity measure, as well as clustering and thus the construction of a typology of such objects require a meaningful, reliable and practical (real-valued) measure of distance between any two such objects. For the distance between two graphs with identical nodes (where only the links between the nodes can differ) and where a link, which can be either directed

or non-directed, is either present or absent, [2] the Hamming Distance is a convenient and useful measure of distance (Hamming 1950). But for cases where the edges are either weighted (the one dimensional, continuous case), or where they are described by multiple parameters, a more meaningful measure of difference should be sensitive to differences in such link attributes.

5   References

Anderson, John R. 2008. *Cognitive Psychology and its Implications*. Worth Publishers; Seventh Edition edition.

Ansolabehere, Stephen and Jones, Philip E. 2010. Constituents' Responses to Congressional Roll-Call Voting. *American Journal of Political Science*, Vol. 54, No. 3

Axelrod, R. 1976. *Structure of decision : the cognitive maps of political elites*. Princeton : Princeton University Press.

Berinsky, Adam J., Huber, Gregory A. and Lenz, Gabriel S. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.coms Mechanical Turk. *Political Analysis* 20:351368.

Bostrom, A. M., Morgan, G., Fischhoff, B. and Daniel Read. 1994. What Do People Know About Global Climate Change? *Risk Analysis* Vol 14. No. 6.

Converse, PE. 1965. The Nature of belief systems in mass publics, In *Ideology and discontent*, ed. Apter, D.E. New York: Free Press.

Chong, Dennis, and James N. Druckman. 2007. Framing Theory. *Annual Review of Political Science*. Vol. 10: 103-126.

Grimmer, Justin, and Gary King. 2011. General Purpose Computer-Assisted Clustering and Conceptualization. *Proceedings of the National Academy of Sciences* Copy at http://j.mp/j4xyav

---

[2]In other words the adjacency matrix consist of only binary (0 or 1) entries.

Hamming, Richard W. (1950), "Error detecting and error correcting codes", *Bell System Technical Journal.* 29 (2): 147160, MR 0035935.

Lewis-Beck, M.S and Stegmaier, M. 2000. Economic Determinants of Electoral Outcomes. *Annual Review of Political Science* Vol 3:183-219.

Lombrozo, T. 2006. The structure and function of explanations. *Trends in Cognitive Sciences*, Vol. 10(10): 464-470.

Maibach, E.W, Leiserowitz, A. Roser-Renouf, C. and Merty, C.K. 2011. Identifying Like-Minded Audiencesfor Global Warming Public Engagement Campaigns: an Audience Segmentation Analysis and Tool Development. PloS One, 6(3): e17571.

Sears, D.O., Lau, R.R., Tyler, T.R., Allen H.M. 1979. Self-interest vs. Symbolic Politics in Policy Attitudes and Presidential Voting. *The American Political Science Review*, Vol. 74(3):670-684.

Aklin, M. and Urpelainen, J. 2013. Debating Clean Energy: Frames, counter frames and audiences. *Global Environmental Change*, In press.

Weitzman, ML. 1992. On Diversity. *The Quarterly Journal of Economics* 107(2): 363-405.

Zaller, J. 1991. Information, Values and Opinion. *The American Political Science Review*, Vol 85(4):1215-1237.