Decision Focused Inference on Networked Probabilistic Systems: with Applications to Food Security

Jim Q. Smith, Martine J. Barons and Manuele Leonelli
Department of Statistics, University of Warwick, UK

KEY WORDS Bayesian inference, Common Knowledge, Decision Support,

Abstract

Bayesian technologies have been progressively applied to larger and larger domains. Here necessarily probablility judgments are made collaboratively and it is rare that one agent owns all probability judgments in the system. So interesting new foundational and methodological issues have arisen associated with the status of inference support by combinations of such judgments. In this paper we review some recent work on Bayesian inference underlying integrated decision support for huge processes. We argue that in a surprising number of such dynamic environments it is in fact *formally justified* to distribute the inference between different panels of experts and then use an agreed framework to knit these separate judgments to properly score different policies. We also briefly report recent progress in applying this new technology to develop an integrating decision support system for local government officials to use when trying to evaluate implications on food poverty of shocks in the food supply chain if various ameliorating policies are applied.

Acknowledgement 1 *This research was supported by EPSRC grants EP/K039628/1 & EP/K007580/1 and Warwick University Food Global Priority Programme.*

1. Introduction

There are now many probabilistic decision support systems for a use in a wide range of environments. These are designed to give benchmark assessments of the efficacy of various types of policy and to evaluate the impacts of shocks to and progressive degradation of the processes being described in the system. Decision support systems are becoming progressively large and often need to use sophisticated architectures and sometimes also sophistciated numerical algorithms to be able to calculate the outputs needed by the user to inform their decisions.

However there are many environments where decisions need to be based on several cascading or parallel and multifaceted stochastic processes. Each component of these systems can be supported by probabilistic models but the sometimes bewildering array of outputs need to be composed to together somehow before a decision center can compare the efficacy of various courses of action open to it. This paper reports some recent methodological developments to support inference in such huge and complex environments. Many of these have been reported in [22], [23], [24] and especially [38] where many of the detailed technical developments used in this report are reported. We then reflect on the promise and future challenges facing us in this field.

One author's first exposure to this problem - in the wake of the Cherobyl disaster - was to work with Simon French and others for RODOS [2], [39] in the development of uncertainty handling within a support tool for a decision's centre's crisis management after a nuclear accident over 25 years ago. Here various components of the description of a threatened developing crisis - probability models of the processes at work within the nuclear plant, probability models of the dispersion models of the contamination, of the absorption of the contamination into water supply and the food chain and several models of health risk given exposure - were all supported by software developed by different panels of experts. The results of these sometimes very complex pieces of software then needed to be presented to the decision centre to support their management of the crisis. A useful summary of how this was achieved can be found in [27].

$$
\begin{tikzcd}
\boxed{\text{plant}} \arrow[r] & \boxed{\text{dispersion}} \arrow[r] \arrow[d] \arrow[dr] & \boxed{\text{food chain}} \arrow[d] & \boxed{\text{demography}} \arrow[dl] \arrow[d] \\
\textbf{Radiation Risk} & \boxed{\text{inhalation}} \arrow[r] & \boxed{\text{exposure}} \arrow[r] & \boxed{\text{health}}
\end{tikzcd}
$$

The architecture behind this earlier development, although sophisticated for its time, was challenged by the prevaling culture. This meant in particular that within some of the more complex components of the system uncertainty associated with forecasts was often not even formally acknowledged - the associated computational demands helping to provide an alibi for this. So any integrating architecture was forced to ignore uncertainty - at least in some sources of the process. In fact, although the online estimation of parameters which was often acknowledged within some of the better components of the system themselves, this uncertainty was not usually transferred into the composite system. So decision makers within the crisis management system were then left to fold in these uncertainties as best they could - aided by some simple heuristics - to arrive at an integrated assessments of the likely efficacy of various policies to address the overall flow of the potential crisis. Statisticians and decision analysts understand how misleading these heuristics can be [22].

However since that time there has been an enormous technological advance in the capability and speed of probabilistic expert systems that form the components of such systems. Advances in Bayesian Networks (BNs) especially Object Orientated ones [20], Multiregression Dynamic Models [31], probabilistic emulators supported by Gaussian Processes,[19] and a variety of other Bayesian spatio/temporal models has meant that when properly tuned - the component probabilistic models can now produce almost instantaneously accurate expectations of arbitary functions - and especially the variances of any conditioning variables needed to property evaluate the efficacy of various different courses of actions. So it is timely to next develop proper inferential methodologies that can harness this information appropriately and use this in a formally appropriate way to guide the evaluation of policies which can take proper account of all the component uncertainites within such a system.

Two years ago we were charged with developing a proper inferential system that would be both formal and feasible to address uncertainty handling in such environments. We have recently reported this work in [22] and [38]. We are now beginning the process of applying this methodology to a new domain. Over this time, whilst fear of the next nuclear accident has waned and the world has become better protected through good countermeasure plans to this threat, there has been a growing awareness of the challenges of food security both locally and globally. This has most recently been stessed by climate change, population explosions and the developing competition for food in second and third world countries which is changing both demand for food and its affordability everywhere else in the world. Consequently there is an imperative within the western world to develop a decision support tool for local government to help them address the various threats of food poverty within their populations. As in the nuclear example above, these types of processes are dynamic and spatial and can be conveniently broken down into a number of separate components each overseen by its own panel of experts. The methodological developments outlined here have been informed by the our experience in the study of nuclear crisis management but developed with this new application in mind.

$$\begin{tikzcd} \boxed{\text{production}} \arrow[r] & \boxed{\text{supply chain}} \arrow[r] \arrow[dr] & \boxed{\text{shop?}} \arrow[d] \arrow[dr] & \boxed{\text{demography}} \arrow[d] \\ \textbf{Food Security} & & \boxed{\text{price}} \arrow[r] & \boxed{\text{health + educ.}} \end{tikzcd}$$

In the next section we present some of the special challenges in adapting foundational statistical thinking so that methodology can be developed that can inform decision support systems for huge systems like the two outlined above. After discarding some obvious solutions as infeasible we propose a different solution based on a new distributed decision focused methodology. Then in Section 3 we report some recent results about when such systems are applicable. In Section 4 we illustrate through a toy example how the system can use algorithms like tower rules to integrate uncertainly in practice and briefly describe how similar methodologies extend to large systems. We conclude by discussing some of the promise and challenges facing this development.

2. Integrating Decision Support

2.1 Some Special Features

Perhaps the most important distinction between the standard setting for Bayesian decision theory and the one encountered in our scenarios is that the decision maker is *a center rather than an individual*. Even when - as in our examples - this center is constituted of individuals who largely want to act constructively and collaboratively to formally capture the underlying processes driving the crisis it is nevertheless necessary to address this multiagent environment as a game. In particular all rationality ideally needs to be expressed through hypotheses that form the common knowledge base of the agent panels.

Taking this on board, a second important distinction is that typically here each agent has expertise only about particular aspects of the problem from which the center needs to draw. Any common knowledge base within this game must therefore capture a formal structure that is able to represent a unanimity about *who might be expert about what*. In particular it needs to capture what it might mean for the different agents to be prepared to adopt the beliefs of the most appropriate domain expert panel. Under such conditions it will then be rational for panels to agree to *delegate* their reasoning and evaluation to the appropriate domain experts. In the next section we outline how a center's probability distribution can be constructed around the salient features of a probability distribution.

Thirdly it will typically be necessary within these environments for a center to be able to *justify its choices* to the outside world and to be able to give a plausible explanation of the reasons behind its choices. This is unlike many single agent systems. There the agent makes the best probability judgments she can - using her own personal and sometimes only partially explicable evaluations - to obtain a good outcome. Furthermore that individual is often also free to choose what "good" might mean in her given context without needing to justify that choice.

A center managing a crisis rarely enjoys this freedom: it will also usually need to be able provide the rationale behind the adoption of a policy to supplement the policy itself. In such a scenario the center will therefore need to be able to provide:

1. An agreed *qualitative structure*, providing a plausible description about how different features of the development relate to one another and how the future might potentially unfold. This structure must be transparent enough to be understood by all experts in the systems.

2. A compelling *narrative* based on best evidence about what might happen within each component of the process.

3. A plausible *numerical evaluation* within each component of the extent to which the critical variables within the system might be affected by the developing environment when the most promising mitigating policies might be applied.

As well as encapsulating all the elements above - which concern the underlying process - any common knowledge base must, of course, also be sufficiently rich to contain an agreed set of policies that might be considered and an appropriate utility structure on which the efficacy of these different options can be scrutinized. Furthermore the Bayesian paradigm demands that it must be possible to calculate the expected utility scores for each potential policy applied to this huge system and to evaluate these policies *accurately and quickly* with respect to a shared probability measure.

Although these challenges appear almost insurmountable, there are in fact certain factors in our favour. The first is that a center with a remit like the ones described above is not usually concerned that the composite system provides auditable and compelling judgments about *everything*. It will typically be responsible for properly delivering and explaining only those aspects of the process that might have a significant impact on the critical features of any unfolding crisis within this remit. Within a Bayesian context these critical features are defined by the attributes of a utility function specified by the center.

Of course such attributes need to be elicited. However this is one of the more straightforward tasks in building support. For example, in the context of evaluating countermeasures after an accidental nuclear release this process was successfully conducted decades ago. There appropriate measures could be categorized into three subsets: measures of the predicted health consequences on the population, the public acceptability of any policy and the resource implications of applying particular policies to a given scenario. Another example is given in our most recent project: through a sequence of decision conferences a local authority have outlined four main categories within which to assess the impact to them of food poverty within their jurisdiction, each measured by a well defined vector of attributes. In our first parse these factors were articulated as the effects of malnutrition or threats of malnutrition on health, the effects on children on their academic performance, the potential for social unrest - such as riots - provoked by the non availability of food stuffs and of course the cost and resource implications of applying any ameliorating strategy.

There is therefore often a strong focus on a small number of measurable consequences associated with an unfolding crisis. Now, of course, the types of description we have in mind must be rich enough to explore the knock on effects that might happen to components of the system when that system is stressed by abrupt changes to the physical environments or new policy directives it might receive. We see later that the progressive impact of such shocks can often be conveniently modelled though chains of causal relationships between the mediating processes when the term "causal" has a precise technical meaning.

Despite the challenges presented by these causal chains it can often be shown that there nevertheless exists a proportionately much smaller vector of variables which might significantly impact on the utility attributes of the problem that would be the case were we using the system to solve completely general inferential tasks. So this vastly reduces the modelling task, gives guidance about the necessary underlying granularity in space time and type of the integrating model and the players whose judgments will be needed in order to score different policies. In particular it is not necessary to capture all available expert judgments for such support but only those features that might be critical in helping to discriminate between the potential effectiveness of one enactable policy against another.

There is a second reason to be optimistic about the feasibility of developing this sort of support. There has been a recent vigorous development of various graphical model classes for example object oriented Bayesian networks and these now enjoy a strong formal foundational basis. These frameworks can provide an overarching structure around which to model processes whose variables can exhibit highly heterogeneous relationships to one another. Now sadly in practice for the scale of the problems we have in mind here there is often no generic framework - and so no generic software - which is either logically capable of faithfully expressing our underlying process or sufficiently focused and powerful to make calculations quickly enough to be of practical use.

However what this development has given us is new *inferential axioms* that provide a way of scrutinizing and justifying in a generic way many different families of models - especially those that can be depicted by different families of graphs. Such axiomatic systems - for example semigraphoid, graphoids and separoids [28],[30], [11], [36] have provided compelling reasoning rules to justify qualitative hypotheses about whether or not one piece of information is relevant to the prediction of a second given information from a third. These are often couched in terms of rules about reasoning about irrelevance. In our context we argue that these reasoning rules can be plausibly accommodated within the common knowledge framework of the multiagent game discussed earlier describing the collaboration of agents in the center. Thus let $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ be arbitrary vectors of measurements in the product space of variables defining the DM's problem.

Definition 2 *Say that the client believes that the measurement $\boldsymbol{X}$ is* irrelevant *for predicting $\boldsymbol{Y}$ given the measurement $\boldsymbol{Z}$ (written $\boldsymbol{Y} \amalg \boldsymbol{X} | \boldsymbol{Z}$) if she believes now that once she learns the value of $\boldsymbol{Z}$ then the measurement $\boldsymbol{X}$ will provide her with no extra useful information with which to predict the value of $\boldsymbol{Y}$.*

We next assume that the centre accepts that for their problem all aspects of dependence satisfy the semigraphoid axioms. Explicitly this means that any irrelevant operator $\amalg$ chosen by the center respects two properties see [36]. The first, called the *symmetry* property, asks that for any three disjoint vectors of measurements $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$:

$$\boldsymbol{X} \amalg \boldsymbol{Y}|\boldsymbol{Z} \Leftrightarrow \boldsymbol{Y} \amalg \boldsymbol{X}|\boldsymbol{Z} \tag{1}$$

This property holds for most probabilistic and non-probabilistic methods of measuring irrelevance, Even more compelling - see e.g. [28]. for an explanation of this - is a second property, called *perfect composition.* This asks that for any four disjoint vectors of measurements $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W}$,

$$\boldsymbol{X} \amalg (\boldsymbol{Y}, \boldsymbol{Z})|\boldsymbol{W} \Leftrightarrow \boldsymbol{X} \amalg \boldsymbol{Y}|(\boldsymbol{W}, \boldsymbol{Z}) \;\&\; \boldsymbol{X} \amalg \boldsymbol{Z}|\boldsymbol{W} \tag{2}$$

Bayesians automatically satisfy this reasoning rule as do a host of alternative inferential systems.

These two reasoning rules together with various statements about relevance within the system at hand together with a finite numbers of other qualitative hypotheses can then be used to populate a common knowledge framework belonging to a decision centre. Note that because the widely used BN models use such reasoning rules, these rules are now well researched and their plausibility are widely accepted as valid.

It is these properties that will allow us formally to appraise when it is or is not appropriate to attempt to systematically integrate judgments for large scale decision support appropriate. This allows us then to *customize* a given center's semantics over a bespoke sets of hypotheses - not necessarily expressible within in a single current generic graphical framework - but nevertheless enjoying the same level of justifiability of more established frameworks. How we proceed to develop such frameworks and how they can be used to guide the inference needed by our centres is described in more detail below.

2.2 Distributivity and the autonomous elements of a supporting narrative

We call a support system which is able to use irrelevance axioms and other agreed structural assumptions to coherently knit together the expert judgments of several different panels with diverse expertise a *integrating decision support system* (IDSS). For such a system to be formal and functional we usually need to be able to prove that the system can perform its task in a distributed way. By this we mean that it is legitimate for each component panel to reason autonomously about the parts of the system over which they have oversight and that the center can then legitimately adopt the delivered judgments of the nominated expert panel as its own. The first reason we need distributivity is that it is usually impractical, inappropriate and often extremely time consuming to demand that panels meet to agree numerical combinations of expert judgments - especially when no-one panel shares good knowledge about the interface of the two areas. A second issue concerns the construction of the narrative we have argued above is likely to be needed to support any policy choice. If the judgments expressed within the system are not consistent with those expressed by the particular panel which is supposedly expert in that domain then how can those judgments be credible?

Thankfully, if an appropriate common knowledge framework is adopted by a center, they ensure that there is no demand which implicitly allows different panels' judgments to contradict one another and the delivery is sufficiently rich ("adequate") for the qualitative common knowledge structure to provide formulae and algorithms to knit together panel quantitative donations to fully score its options, then the semigraphoid axioms enable us to prove that this is possible in a wide range of contexts: see below. This means that it is legitimate for each panel to autonomously populate the system with their own quantitative local domain knowledge, sometimes supported by their own much more detailed dynamic probability models such as Dynamic Bayesian Network [20],Multiregression Dynamic Model, [31] or event tree [36]. As more observational, survey and experimental information becomes available to a particular panel they can then transparently update their beliefs dynamically using these models if necessary and continually refine their inputs to the system without disrupting the agreed overarching structure and its quantitative narrative. Furthermore we will see that when such distributivity is possible it is often the case that each panel need only donate a vector of prearranged conditional expectations for scores to be calculated. This in turn makes it possible to score each policy option almost instantaneously.

3. A Formal Integrating Decision Support System

At this point it is convenient to introduce some terminology Thus we first think of the decision center as a rational expected utility maximizing *SupraBayesian (SB)*. The SB takes the agreed structural framework discussed above. It then embellishes this framework with summaries of some predetermined conditional expectations $\{\Pi_i(d) : d \in D\}$ about various quantities of interest when a policy $d \in D$ might be employed - these expectations being donated by an appropriate panel of experts G_i $i = 1.2, \ldots, m$ - one of the m panels of experts that will inform the integrating system. SB then plans to use these inputs together with the center's common knowledge framework to construct the expectations $\Pi = f(\Pi_1, \Pi_2, \ldots, \Pi_m)$ needed to calculate her expected utilities $\overline{U}(d)$ for each $d \in D$. The plan is then that these scores will be approved and owned by everyone.

But are there circumstances when such a combination is formally justified? The answer is "Yes" surprisingly often. Here is a recent theorem proving one such case. Let $I_0(d)$ be information *common knowledge* to all panels, $I_{ij}(d)$ be information panel i brings to $\boldsymbol{\theta}_j$ $i, j = 1, 2, \ldots, m$, $I^+(d) \triangleq \{I_{ij}(d) : 1 \leq i, j \leq m\}$ and $I(d) \triangleq \{I_{jj}(d) : 1 \leq j \leq m\}$

Definition 3 *An IDSS is* adequate *if SB can calculate* $\overline{U}(d)$ *from delivered outputs,* delegable *if for any* $d \in D$ $\exists$ *a consensus that* $\boldsymbol{\theta} \amalg I^+(d) | I_0(d), I(d)$, *&* separately informed *if* $\amalg_{j=1}^{m}(\boldsymbol{\theta}_j, I_{jj}(d)) | I_0(d)$.

Definition 4 *An IDSS is* sound *if adequate & by adopting the structural consensus all panel members can faithfully adopt* $\overline{U}(d) : d \in D$ *calculated from probabilities donated by relevant panels of domain experts as their own.*

Assuming the semigraphoid axioms above we can then prove the following theorem.

Theorem 5 *An adequate,delegable & separately informed IDSS is sound.*

Proof. See [38] ■

So we have a set of conditions under which an ideal type of IDSS can be built. Furthermore these conditions, whilst not always being satisfied are possible to scrutinized in common language. Through discussing which information sets may or may not be relevant when making inferences about different elements of the multivariate processes the center can determine whether or not a particular framework fulfills the requirements of the theorem above. Note in passing here that this theorem does not only concern probabilistic systems but also any inferential system agreed by the center which satisfies the semigraphoid axioms and which can deliver scores unambiguously - e.g. linear Bayes.

The necessity for adequacy is obvious and the condition of delegability is simply a formalization of the demand that each expert panel is assumed by everyone to be sufficiently well informed to be genuinely more expert than others in the system. The critical assumption is therefore that panels are separately informed. Within a Bayesian context we can use the usual properties of conditional independence to usefully break this condition down into a set of two separate conditions - prior panel independence and likelihood separability - which together are equivalent to the system being separately informed.

Definition 6 *We have* prior panel independence *if* $\amalg_{j=1}^{m}\boldsymbol{\theta}_j, |I_0(d)$. *Data* $\boldsymbol{x}$ *with likelihood* $l(\boldsymbol{\theta}|\boldsymbol{x},d)$ *,*$d \in D$*, is* panel separable *over* $\boldsymbol{\theta}_i$*,* $i = 1, \ldots, m$ *when*

$$l(\boldsymbol{\theta}|\boldsymbol{x},d) = \prod_{i=1}^{m} l_i(\boldsymbol{\theta}_i|\boldsymbol{t}_i(\boldsymbol{x}), d)$$

where $l_i(\boldsymbol{\theta}_i|\boldsymbol{t}_i(\boldsymbol{x}))$ *is a function. of* $\boldsymbol{\theta}$ *only through* $\boldsymbol{\theta}_i$ *and* $\boldsymbol{t}_i(\boldsymbol{x})$ *is a function of the data* $\boldsymbol{x}$*,* $i = 1, 2, 3, \ldots, m$*, for each* $d \in D$*.*

Those with some knowledge of Bayesian inference within BNs will recognize panel independence in that context where different panels can have oversight of different nodes given their parents, as simply a generalization of the global independence assumption. This assumption is almost universally adopted in practical applications of BNs.

The critical assumption therefore is that the collection of data sets gives a likelihood that separates over the subvectors of panel parameters. Of course, this is far from automatic. Even if the system is carefully and compatibly structured it may be impossible to define the parameter vector $\boldsymbol{\theta}$ of the likelihood of a given statistical model in this way - especially in the presence of unobserved confounders. And when vectors of observations can have missing values then this condition is also almost inevitably violated. However there are also many circumstances when this condition can apply. This is most common in asetting where any observational data accommodated into the system is complete and when the underlying dynamic structure is causal in a sense that generalizes the definitions of Pearl [30] so that they can also apply to domain other than the simple BN. We will discuss why causal systems often lead to distributed IDSS below.

When likelihoods are not separable then we can, of course, still approximate - for example using techniques like Variational Bayes. Our formal framework above thengives us a benchmark against which to judge such an approximation. Alternatively - and perhaps more in harmony with the game theoretic basis of this type of analysis - we can instead assume that SB imposes an *admissibility protocol.* This would demand that expert judgments used in the system would only be based on information that would not give rise to ambiguity in subsequent joint inference. Even though it might cause some divergence between public pronouncements made by the IDSS and the private beliefs of panel members, the need for each individual panel to explain its reasoning to outsiders strongly encourages the adoption of such a protocol. Furthermore it has the expedient tendency of being conservative about the accuracy with which various assertions can be made. To adopt such a protocol the center would of course need to agree that only certain types of evidence be accommodated into the system. However note that such protocols - and most notably those of Cochraine Library- are currently widely used within decision support systems designed for collections of users.

3.1 Causal Hypotheses and their relationship to a distributed IDSS

Led by Pearl [30], many authors have recently set about formalizing what is actually meant by causation by framing causal hypotheses in terms of control. All the original work centered on causal hypotheses that could be captured through a BNs. However the semantics have since been extended so that they can also apply to other frameworks. see e.g. [21], [18], [12],[41], [42]. Typically these assume that there is an implicit partial order to the objects in the system that provides the basis of a putative causal order (see eg [33]). Using this partial order we then assume that the joint distributions of variables not downstream of a controlled variable remain unaffected by that control whilst the effect on downstream variables in response to this control of a causal variable to a given value is the same as if the controlled variable had simply taken that value [38]. Many of the newest of these generalizations apply these principles to the sorts of stochastic processes that typically describe an unfolding threat: see [38] for a review of some of these advances.

We saw above that in most of the IDSSs needed to entertain the predicted effects of different potential policies that might be applied to try to control the adverse affects of a threatened crisis. In the reference above we show that if a center adopts various causal hypotheses which exploit the generalizations of "causality" to this dynamic domain then structural hypotheses can be articulated and if appropriate adopted into the common knowledge basis of the center. In this way causal hypotheses help frame the underlying inferential methods.

There is, however, perhaps an even more compelling reason for demanding conditions related to causal hypotheses if an IDSS is to be valid. Note above that when we encourage expert panels to accommodate information into an IDSS, the information they would like to input will often arise from designed experiments. Here, within such experiment, covariates are often controlled to take specific values. The experimenter then assumes that the parameters she estimates in these experiments can be equated with parameters in the observational system defining the development of the crisis. Furthermore she typically assumes that the parameters of observational system will still respect the same probability law as that of the parameters in the experiment. This point was recognized some time ago by Cooper and Yoo [7] who developed collections of assumption which enabled formal learning of discrete BNs where some available data came from designed experiments rather than observational studies. They noted that if the BN was causal in the sense given above then experimental data could be introduced in a simple way. This technology has recently since been extended so that it also applies to other domains (see e.g. [14], [8]). It is interesting to note that, from a methodological point of view, the panel independence assumption which is necessary to ensure distributivity of an IDSS is in fact intimately linked and plausible only when certain causal hypotheses can be entertained: see [10].

However again in many settings such causal hypotheses are plausible - indeed very often made unconsciously see [38]. In particular note that if a panel designs an experirment well then randomization and conditioning often leads to a likelihood which is a function only of its own parameters. So in this case the likelihood trivially separates. And then it follows trivially that the likelihood of any *collections* of such experiments also separates.

So, for example, it can be shown fairly straightforwardly that when there is a consensus that quantitative causal structure is a (dynamic) *causal BN* or *casual Chain event graph* or a *causal multiprocess model* & an IDSS is sound at any time t: then that IDSS remains sound under a likelihood composed of ancestral sampling *experiments* as well as observational sampling: see [38] for examples of such results. It follows that many of the IDSS frameworks we would like to use can be designed so that they are distributive, especially if the center is prepared to entertain the possibility of vetting some of the available evidence as too ambiguous to be formally accommodated into the system. How we can exploit this property is discussed below.

4. Tower Rules and Efficient Transfer of Information

4.1 Assuming the IDSS is distributive

If an IDSS is distributive then it is often possible to prove, provided the agreed form of the utility function has an appropriate polynomial form, that each panel often needs to deliver only a few conditional expectations and not whole joint distributions. This is because the types of structural overarching frameworks embed collections of conditional independences which lead to particular tower rules being respected: see [22]. This in turn means that each panel often needs only deliver a short vector of conditional moments. The SB is then able to evaluate a number of polynomials in these donations recursively to calculate the expected utility scores of the different policies she has available to her. So the various contributions needed from the different panels can be quickly elicited at any time. Furthermore the necessary calculations can be made almost instantaneously. In particular this allows us to hard wire into the IDSS various formulae - looking like forward expectation propagation algorithms [9] that then can be used to make all its necessary calculations for the center. Of course the form of these functions will be customized to the particular underlying framework agreed across the different panels on the underlying.

Once these formulae are in place each panel is encouraged to update its inputs in light of any new information available to it - either concerning the nature of the current unfolding crisis or as it inputs new data from recent experiments and surveys or refines its expert judgments. Note that whenever new data is accommodated this will require the panel to perform a prior to posterior analysis and often in these large environments this will involve performing new numerical analyses. However such numerical analyses routinely and trivially can calculate the numerical values (conditional means, variances) of terms in the conditional moments which the SB needs.

If the plausibility of some of the outputs donated by a particular panel is queried by an outside auditor or another panelist then this request for clarification can be referred to that panel. Because the judgments donated by this panel are its sole responsibility, it can use any current software it owns and documentation of its underlying statistical model class to provide a much more detailed explanation of how its evaluation has been arrived at and why the judgments it expressed are appropriate. This facility is critical to any decision center of the form we discuss here because it may well be that the situation as it dynamically evolves no longer supports some of these background hypotheses. If this happens then this can be quickly fed back to the panel so that it is able to adapt its donations in the light of this new information.

In order to see how this process can be enacted. consider the following toy example.

Example 7 (A Tower Rule for Food) *Consider the following hypothetical framework where the effect of malnutrition on children's educational attainment in a routine region wide state school test of academic ability in a the population of 11 year old children. This is analogous to one of the educational attributes used by a local authority to measure one deleterious impact of food poverty within its catchment. Here we consider only two panels which the center has as common knowledge are currently panel independent. The first, G_1, has taken the various belief inputs it needs from other panels - associated, for example, with predictions of the economic climate that apply in the forecast period, the predicted availability of food in the current crisis and household demography indexed by income and number of children to determine the distribution of an index X of the level of malnutrition across the relevant population of 11 or 12 year olds under study at the time of the next future test The second panel, G_2, is expert in determining the likely SATS performance Y over this population given this index. Various policies $d \in D$ are proposed both aimed at supplementing the diets of this particular group of vulnerable children and in directly enhancing those children's education. Suppose it is commonly agreed that a marginal utility function of this attribute U is an arbitary function of d but is a function of Y as a polynomial of degree no greater than 2. Note that this will then imply that all scores $\overline{U}(d)$ will be expressible as a function of d, $\left\{ m_Y(d) \triangleq E(Y(d)) : d \in D \right\}$ and $\left\{ \sigma_Y^2(d) \triangleq Var(Y(d)) : d \in D \right\}$.*

In this setting the nutritional expert panel $\mathbf{G}_1$ needs only donate

$$\Pi_1 \triangleq \left\{ m_X(d) \triangleq E(X(d)), \sigma_X^2(d) \triangleq Var(X(d)) : d \in D \right\}$$

To predict the performance index Y of these exam results G_2 plans to use the simple Bayes Linear model

$$Y|X, \theta = \theta X + \varepsilon$$

Using this model G_2 is able to calculate

$$\left\{ \mu(d) \triangleq E(\theta|d), \sigma^2(d) \triangleq Var(\theta|d), \tau^2(d) \triangleq Var(\varepsilon|d)) : d \in D \right\}$$

and so the conditional expectations needed by SB which are

$$\Pi_2 \triangleq \left\{ \begin{array}{c} E(Y|X) = \mu(d)X, \\ E(Y^2|X) = \left(\sigma^2(d) + \mu^2(d)\right) X^2 + \tau^2(d) : d \in D \end{array} \right\}$$

Now the standard Tower Rule gives us that for each $d \in D$

$$\begin{aligned} m_Y(d) &= \mu(d) m_x(d) \\ \sigma_Y^2(d) &= \left(\sigma^2(d) m_x^2(d) + \tau^2(d)\right) + \sigma_X^2(d) \left(\mu^2(d) + \sigma^2(d)\right) \end{aligned}$$

So the center can combine the expert judgments of the two panels using these polynomial formulae to calculate the scores it needs. Note that the delivered expert judgments here can be *associated with different levels of complexity. For example G_2's assessments Π_2 could be based on non-conjugate sampling themsleves based on many diverse forms of relevant experiments, in which case it would usually only be possible to deliver numerical values of the required summaries $\{\mu(d), \sigma^2(d), \tau^2(d) : d \in D\}$ and not the formulae behind their calculation. In this example this would not matter. The scores of the competing policy options can still be calculated trivially.*

Now of course this example is absurdly simple. We have suppressed the dynamics of the problem, the fact that the linear models used in these circumstances have a number of covariates, that the population needs to be specified as aggregates of various different subpopulation and that the recurrences range over many such steps. However although such necessary embellishments lead to polynomials of much higher degree and dramatically longer vectors of donations, the form of these polynomials and their calculability nevertheless scales up under very general conditions. The nature and construction of these recurrences, as a function of various types of hypotheses and assumptions, are now well documented and discussed in detail in [22]. In the most complex scenarios these recurrences can be still often be expressed in terms of relationships between high dimensional tensors.

5. Conclusions & Future Research

Currently we are well on the way to building a working integrating decision support system to address issues of food poverty. We have found that most components of the system can be plausibly structured so that each panel component is distributed. We are beginning to discover that quite decent support can be given on the basis of a rather small scale digest of the processes with a total of a few hundred inputs needed from our panels where some of these inputs can be supplied in very routine and transparent ways. So, at least within this domain, the integration of the various probability distributions is feasible and the supporting evaluation can be made to be very quick. We are finding that the impact of judgements at the end of the chains have the biggest impact on the scoring and so currently we are initially concentrating on eliciting and modelling these. Rather interestingly the elicited attributes seem to have close parallels to areas of responsibility that have been independently defined by various local government councils. When, as here, the assessment of attributes already has an obvious owner the elicitation of the judgments and utilities is obviously much more straightforward.

The inferential foundational issues on which this paper focuses appear particularly interesting. Firstly we see here that although the analyses we present here are designed to be shared by many agents and observers the system is nevertheless in no sense "objective". To set up uninformative priors and "let the data speak for itself" is clearly impossible in this type of setting: so much strong domain knowledge and so many domain judgments would be needed before anything sensible could be delivered by the IDSS. What we can build is instead a system based on a kind of *benchmark subjectivity* which captures all that can be said unambiguously: the agreed common knowledge structure collated together with all admitted supporting information form the different expert panels. It represents the *expressed* shared judgment of all participants when they trust one anther's particular expertise and need to be able to be confident that they can justify their choices. We would argue in fact that this is perhaps more useful than something labelled as objective and has interesting links to Smet's ideas [35] of pignistic probability: a gathering of assessments based on what can be agreed before discussions and divergence of opinions take place. Note, in particular, that casting inference in terms of decision support places people rather than outputs from hard wired algorithms at the center of the decision making process which we believe is most appropriate to inference within large systems: essentially seeing this as an activity best addressed by applied statisticians rather than machine learners. In what we describe above probability model outputs have a vital but secondary role to the underlying decision making processes.

Secondly the sorts of formalisms we introduce here need not be conventionally Bayesian. Any reasoning system which satisfies the semigraphoid axioms has the potential for providing the basis of an IDSS of the type we discuss above. The main reason we have focused on probabilistic systems here is simply because these are widespread and have been demonstrated to be provenly useful over a wide range of application. However other methods based on belief functions or linear Bayes methods [17] could, perhaps, prove even more efficacious. The latter option might be especially attractive because it would allow further simplifications of the inferential structure. Only transparently justified statements would then be used within the different panel's accommodation of information. We are currently exploring the efficacy of such methods.

Finally because we have discovered that collection of polynomial equations so often describe the embellished structure of an IDSS, it appears that often techniques using computer algebra [3],[5] provide an especially useful framework for determining the donations needed by the different panels. Techniques borrowed from algebraic and differential geometry can be applied both to construct bespoke efficient algorithms for quickly computing the scores the center might need in huge systems and also for formally studying the robustness of evaluations to various types of perturbations. For some recent initial work in this area see [25] and [26].

REFERENCES

Barons, M.J. Zhong,X & Smith,J.Q. (2014) "A Dynamic Bayesian Network for decision support for sugar food security" CRiSM Rep. (submitted)

Caminada, G., French, S., Politis, K. and Smith, J.Q. (1999) “Uncertainty in RODOS” Doc. RODOS(B) RP(94) 05

Capani, A., Niesi, G., Robbiano, L., 2000. "CoCoA 4.0, a system for doing Computations in Commutative Algebra". Available at cocoa.dima.unige.it.

Costa L, Smith,J.Q. and Nicholls, T. and Cussens J. (2015) "Searching Multiregression Dynamic Models of Resting-State fMRI Networks Using Integer Programming" Bayesian Analysis Vol. 10, No. 2, 441-478.

Char, B., Geddes, K., Gonnet, G., Leong, B., Monogan, M., 1991. "MAPLE V Library Reference Manual". Springer-Verlag, New York.

Clemen, R.T. and Winkler , R.L (2007) "Aggregating Probability Distributions" .in Advances in Decision Analysis Cambridge University Press 154 -176.

Cooper, G.Y. & Yoo, C "Causal Discovery froom a mixture of Experimental and Observational Data" Proceedings of the 15th Conference in Uncertainty in Artificial Intelligence 116 - 125.

Cowell, R.G. and Smith , J.Q. (2014) "Causal discovery through MAP selection of stratified chain event graphs" Electronic J of Statistics vol.8, 965 - 997.

R.G.Cowell,A.P. Dawid,S.L. Lauritzen and D.J.Spiegelhalter (1999) "Probabilistic Networks and Expert Systems" Springer.

Daneshkhah, A. & Smith, J.Q.(2004) "Multicausal prior families, Randomization & Essential Graphs" Adv. in BNs, 1-17.

Dawid, A.P. (2001) "Separoids: A mathematical framework for conditional independence and irrelevance", Annals of Mathematics and Artificial Intelligence,32, 335 -372.

Eichler,M. & Didelez, V. (2010) "On Granger causality and the effect of interventions in time series" Lifetime data analysis 16 (1), 3-32.

Faria, A.E. & Smith, J.Q. (1997) "Conditionally externally Bayesian pooling operators in chain graphs", Ann Stats. 25,4,1740-1761.

Freeman, G. and Smith, J.Q. (2011) " Dynamic Staged Trees for Discrete Multivariate Time Series: Forecasting, Model Selection and Causal Analysis", Bayesian Analysis, Vol. 06, No. 02, 279 - 306.

Glymour, D., Cooper, G.F., 1999. "Computation, Causation, and Discovery". MIT Press, Cambridge.

Goldstein, M. & O'Hagan (1996)."Bayes linear sufficiency & systems of expert posterior assessments"J. R. Statist. Soc. B 58,301 -316

Goldstein, M. and Wooff, D. (2007) "Bayesian Linear Statistic: Theory and Methods" Wiley.

Harbord, R. M., Didelez, V., Palmer, T. M., Meng, S., Sterne, J. A. C. & Sheehan, N. A. (Mar 2013) "Severity of bias of a simple estimator of the causal odds ratio in Mendelian randomization studies" Statistics in Medicine. 32, 7, 246-1258

Kennedy, M. and O'Hagan, A. (2001). "Bayesian calibration of computer models (with discussion)". J. R. Statist. Soc., B. 63, 425-464.

Korb K.B. and Nicholson, A.E. "Bayesian Artificial Intelligence" CRC Press.

Lauritzen S, and Richardson T.L, (2002) "Chain graph models and their causal interpretations"J. R. Statist. Soc.B ;2002 64t3, 321−361.

Leonelli. M.& Smith J.Q. (2015) "Bayesian Decision Support for Complex Systems with Many Distribtued Experts " Annals of Operations Research 10.1007/s10479-015-1957-7.

Leonelli. M.& Smith J.Q. (2014a) "Dynamic Uncertainty Handling for Coherent Decision Making in Nuclear Emergency Response" Proc. of 2013 ANS, Washington

Leonelli. M.& Smith J.Q. (2013) "Using Graphical Models & Multi-attribute Utility Theory for Probabilistic Uncertainty Handling in Large Systems" IEEE 29th ICDEW,181-192.

Leonelli, M., Riccomagno, E & Smith, J.Q.(2015) "The algebra of integrated partial belief systems" CRiSM Res Rep.

Leonelli, M. Roccomagno E. and Smith, J.Q. (2015) "Using computer algebra to symbolically evaluate discrete influence diagrams" CRiSM Res. Rep.

Papamichail, K.N. and French, S. (2005) "Design and evaluation of an intelligent decision support system for nuclear emergencies" Decision Support Systems, 41,1, 84 - 111.

Pearl,J. (1988) "Probabilistic Reasoning in Intelligent Systems" San Mateo: Morgan Kauffman.

Pearl, J., (1995) "Causal diagrams for empirical research". Biometrika, 82:669–710.

Pearl, J.,(2000). "Causality. Models, reasoning, and inference". Cambridge University Press,

Cambridge.
Queen, C.M., & Smith, J.Q. (1993). "Multi-regression dynamic models". Journal of the Royal Statistic Soc. Series B, 55, 4,849-870.
Riccomagno, E. and Smith, J.Q. (2009) "The Geometry of Causal Probability Trees that are Algebraically Constrained" in "Optimal Design and Related Areas in Optimization and Statistics" Eds L. Pronzato and A.Zhigljavsky, Springer 131-152
Riccomagno, E.M. and Smith, J.Q. (2004) "Identifying a cause in models which are not simple Bayesian networks" Proceedings of IMPU, Perugia July 04, 1315-22.
G. Shafer (2003). "The Art of Causal Conjecture". Cambridge, MA, MIT Press.
Smets P. (2005) "Decision Making in the TBM: the necessity of the pignistic transformation." I.J. Approximate Reasoning, 38, 133-147 (2005).
Smith, J.Q. (2010) "Bayesian Decision Analysis" Cambridge University Press
J.Q. Smith and P.E.Anderson (2007). "Conditional independence and Chain event graphs"Artificial Intelligence, 172, 1, 42 - 68.
Smith, J.Q, M.J. Leonelli, M & Barons, M.J. (2015) "Bayesian Inference in Distributed Expert Systems" CRiSM Res. Rep.
Smith, J.Q.et al (1997). “Probabilistic Data Assimilation with RODOS”. Radiation Protection Dosimetry,73, 57-59.
P. Spirtes, C. Glymour and R. Scheines (1993). "Causation, Prediction, and Search". Springer-Verlag, New York.
Thwaites, P. (2013). "Causal identifiability via chain event graphs." Artificial Intelligence 291–315. MR3024205.
Thwaites, P., Smith, J. Q. and Riccomagno, E. (2010). "Causal analysis with chain event graphs". Artificial Intelligence 174 889–909. MR2722255.