Eliciting Probabilistic Judgements for Integrating Decision Support Systems

Martine J. Barons, Sophia K. Wright and Jim Q. Smith,

Abstract In facing extremely large and interconnected systems, decision-makers must often combine evidence obtained from multiple expert domains, each informed by a distinct panel of experts. To guide this combination so it takes place in a coherent manner, we need an integrating decision support system (IDSS). This enables the user to calculate the subjective expected utility scores of candidate policies as well as providing a framework for incorporating measures of uncertainty into the system. Throughout this chapter we justify and describe the use of IDSS models and how this procedure is being implemented to inform decision-making for policies impacting food poverty within the UK. In particular, we provide specific details of this elicitation process when the overarching framework of the IDSS is a dynamic Bayesian Network (DBN).

1 Introduction

In our increasingly interconnected world, large systems are becoming more common and progressively more complex. This means that statistical modelling protocols must also evolve to accommodate these changes. Typically, in this new situation, decision-makers need to gather evidence from a variety of different expert domains. Each such domain has a limited number of people who are deemed experts in par-

Martine J. Barons
University of Warwick, Coventry, CV4 7AL, UK, Grant EP/K007580/1, e-mail: Martine.Barons@warwick.ac.uk

Sophia K. Wright
University of Warwick, Coventry, CV4 7AL, UK, Grant EP/L016710/1, e-mail: S.K.Wright@warwick.ac.uk

Jim Q. Smith
University of Warwick, Coventry, CV4 7AL, UK and Alan Turing Institute e-mail: J.Q.Smith@warwick.ac.uk

Paper No. 16-05, www.warwick.ac.uk/go/crism

ticular aspects of the interdependent system. So in such systems, we must develop ways to combine together evidence from these different domains into a coherent whole. The evidence we want to accommodate will typically be framed probabilistically and will often be supported by domain-specific probabilistic predictive models. Fortunately, there is now a technique which makes possible coherent inference over a network of these multifaceted probabilistic systems that can provide such decision support for policymakers. We call this composition an integrating decision support system (IDSS), see [41, 42].

1.1 A Probabilistic IDSS: its genesis and functionality

One complication facing the coherent combination of judgements within an IDSS is that, in the 21st century users are now typically teams, here called *decision centres*, rather than individuals. The implicit (albeit virtual) owner of beliefs expressed by this team will henceforth be referred to as the *supraBayesian* (SB). This SB embodies the beliefs of the decision centre. Through this construction we are able to address issues such as statistical coherence or rationality as it applies to the system as a whole.

Once a coherent system has been built, being probabilistic in nature, these huge composite models enjoy many of the advantages seen in probabilistic models of smaller systems. In particular, the algorithms to determine the efficacy scores are based on widely accepted formulae. Furthermore, these algorithms permit the smooth combination of expert judgements with any information obtained from experimental or survey data that might be available to the centre.

The need for such integrated systems became clear to one author of this chapter, whilst working with Simon French and others as part of a team designing a decision support system for operations crisis control after an accidental release of radiation from a nuclear plant. This research was organised as part of a large EU programme called RODOS (Real-time On-line DecisiOn Support) [5, 14]. Part of the decision support led to fast evaluations of the effectiveness of various candidate countermeasures, integrating information from a variety of sources in what was then called an "evaluation subsystem". When addressing this massive problem, we were forced to separate the description of the unfolding processes and their threats into components where, just as we described above, each component was informed by a separate panel of experts. Each panel was charged with providing its own domain information which was then delivered to the centre and combined with the results from other panels to score the efficacy of the different candidate countermeasures. To process information in this way, once the prescribed inputs from other components were delivered, each component would need to be able to take and then produce its outputs *autonomously*.

Although the RODOS development was seminal for its time and the long collaboration produced a valuable decision support engine, urgencies in its development meant that the architects and methodologies on which RODOS depended were often necessarily naïve. Furthermore, due to the constraints on computational availability, such a composite system was constrained to fairly coarse scales.

The project raised some important questions about the construction of decision support systems for multi-faceted problems like this. In particular there were two main concerns:

1. Firstly, this early system was unable to suitably allow uncertainties to be incorporated into the combined assessment in a sound or comprehensive manner. This meant that for example, if the expected consequences of one policy were marginally better than another, but the uncertainties arising from the first policy were much greater than the second, then suggesting to the decision centre that the first option was better than the second was often not formally correct and it may also have had disastrous consequences.
2. Secondly, the actual dynamics of the situation were often only partially incorporated into the evaluation of the system, meaning that it could only provide snapshots of how scenarios were unfolding.

Now, at last, general methodologies are in place that can be implemented to properly process the dynamics of the system [23, 24]. We describe in this chapter how such an implementation is now being enacted. The theoretical development in [41] needed to address two fundamental questions:

1. "Is it even theoretically and logically justifiable to compose inferential methodologies using an overarching system like the one used for RODOS?".
2. "Can the scores associated with these formally correct and justifiable support systems be structured in such a way that the necessary calculations can be made quickly enough for the IDSS to be feasible?".

The first of these two questions, regarding the logical justification of creating such overarching systems, can be broken down into more specific queries: "What conditions guarantee all uncertainties expressed by experts are appropriately represented and processed in such an integrating system? Can the dynamic nature of such processes be captured and the progressively fine-grained information be processed appropriately? Can these, at least in principle, be constructed to guide the evaluation of policies in a way that takes proper account of all the component uncertainties and dynamics within such a composite system?". In order to answer these questions fully it was necessary to show that it was possible to break up a composite DSS into different, autonomously updated components and subsequently aggregate them together in a formally justifiable way.

Recently, we have been able to show that under commonly satisfied conditions, and with a careful construction of the components of the system and their interfaces, expected utility scores of candidate policies can be either perfectly calculated in this distributed way or approximately so. The proofs of these recent discoveries are necessarily rather mathematical and so beyond the scope of this book. However, these are presented in the public domain formally appearing in [41] and less formally in [42] plus [23, 24, 25]. Furthermore, we can show that within such an integrated system the rationale behind the different component forecasts can be delegated to panels overseeing the corresponding components delivering these outputs. Here, we simply present the relevant summary results of this technical work.

In addition to these methodological advances, the second question raised the issue of computational feasibility in relation to speed of such systems for the IDSS to be feasible. Recent results, show the answer to this question also to be "Yes". The theory for this assertion, as it applies to various types of such systems is published in [23]. Therefore, the methodological obstacles that had faced both the justifiability of the RODOS IDSS and its ability to quickly and effectively calculate estimates and their associated uncertainties, have now been surmounted.

We note that both the theory and the algorithms referred to above apply to probabilistic models where both the composite model and its components have associated probability distributions. Here the changing, probabilistically expressed beliefs of the expert panels need to be elicited and then processed. Precisely because these systems are probabilistic, the relevant uncertainties associated with different candidate policies can be expressed directly via distributions and the stochastic dynamics represented in terms of various stochastic processes, using very well understood methods of uncertainty handling.

Of course it is one thing to demonstrate that this type of IDSS is *formally* justifiable and feasible to implement *in principle*, and quite another to apply the proposed methodology to construct an *actual working system* that can be used successfully to help inform a particular domain. Over the last couple of years we have begun to build such a system for UK food security in collaboration with policy-makers, using the theoretical development described above. This chapter describes how we are currently implementing the methodologies and the practical challenges we face when doing this.

1.2 The running example of Food Security

High on the UK government agenda is the issue of food poverty. The world population reached 7.3 billion as of mid-2015, which is the result of an expansion of approximately one billion people in the last twelve years and projections expect the total to pass 9 billion by 2050 [43]. It is therefore vital that optimal use is made of the world's finite resources for food production [6]. With such a growth in population, the demand for food and its affordability has changed everywhere in the world and stresses have been exacerbated by an increasingly extreme division of wealth between the rich and the poor, within and across nations, and the emergence of food riots in 2008 and 2011 [22]. This has resulted in making food poverty endemic worldwide and an increasingly serious threat to even wealthy nations such as the UK. As a consequence, it is very timely to develop decision support tools designed to help government policy-makers to assess the effects of policies they might enact to address the various threats of food poverty. Strains felt by local government in the UK are currently severe in the UK because of progressively decreasing budgets which requires them to implement different forms of cutbacks to financial benefits and spending on social support, within the populations for whom they are responsible.

In this context, local government need an IDSS. The overarching food security IDSS model integrates together all social, political and socioeconomic factors which may affect food poverty within the UK, shown in the schematic below, see Figure 1.

It is possible to use a variety of different overarching frameworks to embody this integration. Here, we shall focus our attention on how we perform this integration when the overarching framework is a dynamic Bayesian network (DBN). Probabilistic decision support systems (DSSs) based upon Bayesian network technologies are now widely used in a variety of environments.

Dynamic Bayesian networks (DBNs) have already proven to be particularly useful when we have multifaceted, interdependent systems which need to incorporate multiple models, natural processes and contextual information. As well as modelling the expected course of a process, these networks can be used as a framework to intelligently integrate uncertainties about the impact of different events which would result from selecting different policy choices, driving mechanisms or external shocks. Furthermore, this framework can be embellished into a full probabilistic model to enable the decision centre to combine expert judgement with data that tracks the unfolding process, as well as utilising new experimental evidence as and when this arrives. These methods use general probabilistic machinery, such as Bayes Rule, in ways discussed in previous chapters of this book.

Whilst developing the IDSS we have found that, perhaps rather predictably, the expert panels largely mirror panels already created by organisations and governmental agencies. In the schematic, each colour depicts a different expert panel. The model is informed through a number of sources, for example:

- Demography and socio-economic status statistics (SES) distributions are available from the Office for National Statistics (ONS),
- Costs of housing, energy and general cost of living (CoL) are available via the consumer prices index (CPI),
- Food trade (imports and exports) and farming yields can be obtained from DEFRA,
- Supply chain disruption and overall food availability can be obtained through DEFRA, the foods standards agency and the food and drink federation,
- Access to credit is available from the Bank of England,
- Household disposable income distributions come from ONS,
- Food costs are via the cost of a typical basket of food, which is systematically calculated as part of the UK CPI calculated by the ONS.

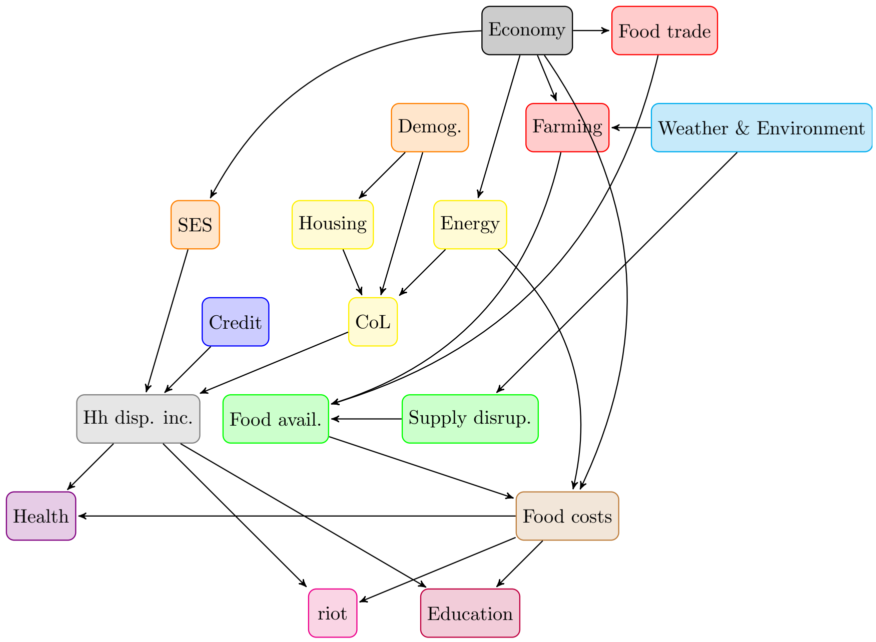

Fig. 1 A plausible schematic of information flows for the modules of a UK food security IDSS. KEY: Economy - UK economic forecasts; Demog - Demography; Farming - food production; SES - Socio-economic status; Credit - access to credit; CoL - cost of living; Food Avail - Food availability; Supply disrupt - food supply disruption; Hh disp inc. - household disposable income.

Within each of these expert panels lies a complex sub-network. For example, consumer price index (CPI) describes a typical basket of goods and services which are purchased by an average UK household and the average price of such a basket is reviewed both monthly and yearly. As we are interested in variables such as food costs, we can use the basket of goods as a measurable proxy. The model of the basket of food needs to be disaggregated into various pre-defined food subgroups such as meat, fruit, vegetables etc. By probabilistically modelling each of these diverse subgroups individually and then reflecting on natural dependencies between them we can derive sub-networks which can be combined together to provide forecasts for the cost of the entire food basket over time.

Another sub-network is concerned with modelling pollination of food crops within the UK and this will form our detailed example in Section 3.2.5. It is estimated that 70% of important food crops are pollinated by bees [8] and the contribution of other insect pollinators is also significant [36]. So the status of food pollinators, and of bees in particular, is a key concern in global food security [3, 27].

In this chapter we will describe in some detail how we constructed the probabilistic component relating to pollination within this system, itself a DBN. We will discuss the implications of the sort of elicitation required within this context and various sorts of diagnostics that can be used to check its plausibility.

2 Framing a Complex Dynamic System

Although IDSSs may seem to be a suitable approach for complex dynamical models, it is not always appropriate to formulate a problem in such a manner. There are a number of criteria which must be fulfilled for an IDSS to be fully justifiable. Building an IDSS for a multifaceted, interdependent system requires the following properties to hold:

1. The decision maker is typically *a centre rather than an individual*. So a necessary condition is that the centre needs to be populated by individuals who largely want to act constructively and collaboratively. In other words, the 'centre' must be motivated to strive to act as a single coherent unit for a common goal.
2. In particular, there must be consensus about the appropriate *utility structure* on which the efficacy of the candidate policies could be scrutinized, were certain unfoldings of events to be certain. For example, such a consensus might be that the centre's utility functions should have preferentially independent attributes.
3. Consensus also needs to be attained about an overarching description of the dynamics driving the process. This can be allowed to take a variety of forms depending on the context. In this chapter, we have assumed that this overarching structure can be represented by a DBN (defined in Section 3.4.1) although other frameworks can be used, for example see [23].
4. The necessary coherence of the group requires there to be a consensus about *who is expert about what* in order to identify appropriate expert panels. In a formal sense, this consensus in turn implies that individuals outside a domain should be prepared to adopt the beliefs of the expert panel of that domain as their own, see [41]. It can then be proved that they should then delegate their reasoning about every domain to the appropriate panel.

Within our two contexts, the overarching food poverty system and the pollinator abundance sub-network, the first condition of a collaborative decision making group was broadly met. This means that all local government officers agreed that food poverty is issue requiring action, although meeting this need impinged on several budget-holding departments hence requiring negotiation and co-operations. Note however that this co-operative approach is not always taken. Many decision centres work more like a court of law, especially those in public health policy making. In these cases, advocates fight both the case for and then against a policy and win the argument in a competitive way. In such a context, the support system we define here can then at best only support one side of the argument. We note however that the sorts of methods we describe here have been used for exactly this purpose in a court of law [34].

The second condition, utility consensus, is often delivered through decision conferencing. Very briefly, a facilitated discussion encourages the centre to first determine sets of broad objectives that preferences of the centre should bear in mind when assessing the efficacy of different candidate policies. Then agreement is negotiated about a vector of attributes of the utility function that best capture the essence of these objectives. These vectors of attributes of the utility function must be specific enough to be measured in an unambiguous way, i.e. passing the clarity test, [17]. Thus, for example, an objective described as minimising "the number whose quality of diet was low over the coming year" could not be treated as an attribute, while on the other hand "the number of individuals treated in hospitals A for conditions which included explicit mention of malnutrition within a region B from Jan 1st 2020 - 2021" would be a candidate attribute.

Within a region the vector of attributes of the utility function then provides not only a transparent and unambiguous picture of what might happen in the future, but also one that balances different aspects implied by a given objective. In the context of the types of decision support we give here it is very common for the utility function to depend on attributes that estimate the impacts of different candidate policies well into the future.

The quantitative form of a utility function will eventually need to be elicited in a manner which is appropriate to the centre. This utility is a function of the possible set of future values of the whole composite vector of attributes. There are many ways to do this that are described in detail in [11, 13].

The next step in the elicitation process, needs only to have identified the attributes of the utility functions but not the quantitative form of the utility function itself. So although discussion needs to draw out those clearly specified measures on which preferences might depend, at this stage it is unnecessary, and indeed often unwise, to assign values of these attributes. This is often most efficiently performed once the overall structure of the problem and its relationships to other processes has been at least provisionally mapped out in a way we will describe below.

Remark 1. The elicitation of an IDSS can only be done robustly if performed in an *iterative* manner, meaning we review the qualitative structure of the IDSS repeatedly. At periodic intervals, previous steps of the process are reviewed and checked in light of the more profound understanding of the process acquired through further elicitation. This enables modification and improvement to the various contributing elements defined in the construction of the qualitative form of the model. The process continues until the decision centre is content that the structure is fit for purpose, called "requisite" [32].

Remark 2. Since the process of elicitation is an iterative one, it is often wise to begin with some simple measures, proceed with an initial structural elicitation, and then to revisit the initial list of attributes of the utility to consider whether these need to be adapted or supplemented so as to better measure the efficacy of the possible candidate policies. In our experience it is often only after the science, economics or sociology driving the process has been more fully discussed that the decision centre can become fully aware of the suitability of certain types of utility attribute measure.

Since household food security (food poverty) is not directly measured in the UK, it was necessary for the local government decision-makers to devise a suitable proxy for this measure, relevant to their areas of responsibility and which are measurable. Through a sequence of decision conferences, we elicited from them three areas which they expected would be impacted by increasing household food insecurity within their jurisdiction: educational attainment, health (as the effects of malnutrition, or threats of malnutrition, on health in the short medium and long term), and social cohesion. Finally, of course, the cost and resource implications of applying any ameliorating strategy to address these negative consequences must be taken into account. Discussions also determined suitable measurements of these four attributes or proxies which would make suitable surrogates.

In the case of educational attainment, it is well-established that on average, pupils who are entitled to free school meals have a lower educational achievement in public examinations than pupils not entitled to free school meals. Eligibility for free school meals is a proxy measurement for deprivation since it is based upon household income. Educational attainment is assessed by a vector of different examination results: students are expected to gain SATs level 4 or above at age 11, and five or more GCSEs, including English and Maths, at grade C or higher at age 16. Department for Education defines disadvantaged pupils as any child in care of the local authority, or any pupil who has been eligible for free school meals at any time over the last six years. Pupils classified as disadvantaged have a lower average educational attainment record than other pupils and there is a direct correlation between level of qualification and employment and earnings.

Health concerns are captured by admission to hospital with a primary or secondary diagnosis of malnutrition, and death with malnutrition listed as a contributory cause. Figures are available from the Hospital Episode statistics for records of diagnoses of malnutrition and death records indicate whether malnutrition was a contributing factor

Social discontent might be expressed in terms of, for example, food riots provoked by the inaccessibility of food stuffs. Food riots are a rare, but costly occurrence, which the decision-maker is keen to avoid. Crime statistics data collected by the ONS record civil unrest including riots, criminal damage and looting.

Cost is measured directly by the amount of money that local councils must spend to implement and uphold any change in policy. For example, estimated costs associated with different intensities of disturbance can be estimated from the resource implications and damage to infrastructure in past riots.

Remark 3. Although the initial problem may look overwhelmingly complex, by focusing the centre and its expert panels on those issues that really impact on final outcomes we can vastly reduce the scope of deliberation. In particular, it soon becomes apparent to all that it is not necessary to capture all available expert judgements for decision support, but only those features that might be critical in helping to discriminate between the potential effectiveness of one candidate policy against another.

Once attributes of the utility have been decided, the next task is to construct a description of the processes leading to the different values of the utility attribute vector. This enables the centre to capture qualitatively how the system might respond to key unfoldings of events. Although this issue has not yet been addressed within this book, our experience, this part of the elicitation process is a critical part of the elicitation process is a critical part of the modelling of large systems. Fortunately, many useful techniques for capturing these explanations have been widely documented and tested against thousands of applications [11]. In the next sections we review the basis of this work. We then describe how we are currently applying these methodologies to construct an IDSS for policies related to household food poverty and the sub-module within that IDSS pertaining to insect pollination of food crops.

3 An agreed picture of the whole probability process

We shall now discuss in detail the process of creating and populating the IDSS. The steps below are based upon published theory and experience of applying these models to specific scenarios. We conclude this section with a look at our food security application.

3.1 An overarching structure and Common Language

In Remark 1, we required the centre to agree an overarching qualitative structure to provide a plausible description about how different features of the development relate to one another and how the future might potentially unfold. Obviously this description needs to be transparent enough to be understood by all experts in the system. In this application it can be expressed as a graph of vertices and edges which together express how variables in the system are believed to relate to one another. The logical foundation and probabilistic compatibility of our description ensures the graph can be expanded later into a full description of the stochastic process driving the the utility attributes in any given unfolding of events. This in turn will enable the centre to evaluate the expected utility scores associated with those policy choices open to it.

This structure needs to be elicited and this elicitation should ideally include representatives of all domain experts across the system as a whole, and is best conducted using common language (as far as is possible). It follows that the type of elicitation we use for this stage is most often behavioural: the experts discuss various options and try to arrive at a consensus about the possible dependences, see [21, 40].

Remark 4. If there is strong disagreement about whether or not a dependency exists in the system then the group of domain experts should assume initially that a dependency does exist and only later explore whether in the face of evidence such dependence is supported. On the other hand, if there is a broad consensus that if a dependence might exist it will be weak then it is usually wise to omit this dependence in the description and only revisit and re-examine this assumption later when the understanding of the underlying process is more mature. In this way we are often able to contain the structure to a manageable size.

The question remains about how exactly we can capture such dependence hypotheses formally without first eliciting probabilities. Fortunately, a mature industry of graphical modelling has now provided us with a new set of inferential axiomatic systems. One language is based on the semigraphoid axioms [9, 30, 31, 40]. These rules simply specify two simple properties we might expect to hold whenever someone asserts that knowing one collection of measurements $\mathbf{Z}$ does not help the prediction of a vector of measurements $\mathbf{Y}$ once the vector of measurements $\mathbf{X}$ is known. More formally and precisely:

Definition 1. Suppose that the client believes that the measurement $\mathbf{X}$ is *irrelevant* for predicting $\mathbf{Y}$ given the measurement $\mathbf{Z}$ (written $\mathbf{Y} \perp\!\!\!\perp \mathbf{X}|\mathbf{Z}$) so that once she learns the value of $\mathbf{Z}$ then the measurement $\mathbf{X}$ will provide her with no *extra* useful information with which to predict the value of $\mathbf{Y}$.

We assume that all in the decision centre accept that when they make an irrelevance statement like the one above then two properties hold [40]. The first, called the *symmetry* property, asks that for any three disjoint vectors of measurements $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$:

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \Leftrightarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{X}|\mathbf{Z}.$$

In words this states that if $\mathbf{Y}$ is irrelevant for predicting $\mathbf{X}$ once $\mathbf{Z}$ is known then also $\mathbf{X}$ is irrelevant for predicting $\mathbf{Y}$ once $\mathbf{Z}$ is known. It is simple to check that this property holds for most probabilistic and many non-probabilistic methods of measuring irrelevance.

Even more compelling is a second property, called *perfect composition*, see for example [30], for an explanation of this. This property asks that for any four disjoint vectors of measurements $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$,;

$$\mathbf{X} \perp\!\!\!\perp (\mathbf{Y}, \mathbf{Z})|\mathbf{W} \Leftrightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y}|(\mathbf{W}, \mathbf{Z}) \;\&\; \mathbf{X} \perp\!\!\!\perp \mathbf{Z}|\mathbf{W}.$$

More informally, this means that when forecasting $\mathbf{X}$ already knowing the value of $\mathbf{W}$, the statement $(\mathbf{Y}, \mathbf{Z})$ are irrelevant to $\mathbf{X}$ is the same as saying $\mathbf{Z}$ is irrelevant to $\mathbf{X}$ and that after learning the value of $\mathbf{Z}$, $\mathbf{Y}$ provides no relevant information about $\mathbf{X}$ either. Bayesian inference automatically satisfies this reasoning rule (as do many other alternative inferential systems).

These two reasoning rules allow us to elicit a collection of irrelevance statements from experts and deduce many others as logical consequences. More importantly, we note that this can all be done without mentioning probability and therefore can be done using common language. Furthermore, the graphs of the BN can be drawn *directly* from these types of irrelevance statements, requiring only one assertion for each vertex in the graph. The BN so constructed has a logical integrity and can be used as an overarching framework based on verbal statements from experts that can then be later embellished into full probability models. Each irrelevance statement of the form above then just transforms into a conditional independence statement in that probability model.

3.2 Defining the features and variables in a problem

In any decision analysis we need to determine which parts of the process are both intrinsic to the description of the process, and are as yet uncertain: either due to lack of hard data in a certain element of the process or because the knock on effects from the implementation of a specific policy are not fully predictable.

3.2.1 What are the centre's attributes and time frames?

We first import into the discussion the preliminary vector of utility attributes of the problem that might inform a one step ahead prediction of the value of each utility attribute vector, [40]. These are termed the "Level 1" vectors, see Figure 2. Within this early phase, clients need to decide what time step is the most natural one to use for the purposes of the support of the IDSS. The appropriate choice of these steps depends on a number of factors: for example the speed of the process, how relevant data is routinely collected on some of the components, and some technical acyclicity assumptions that are typically known only to the analysts. In our application, the agricultural cycle is typically annual whilst pollination is required only for specific months of the year and pollinator lifecycles are measured in weeks. However, the abundance of honey bees is measured in an annual survey. The most crucial factor in determining the most appropriate step length is the timing in which the success of the system will be appraised. This early stage of the elicitation is perhaps the most delicate because of the elicitation because there is often conflict between the granularity of for example, informing economic models of the process and sample survey regularity, and the needs of the system. Typically, we would advise that the needs of the centre and the time horizons at which they work should be prioritised when making this trade-off.

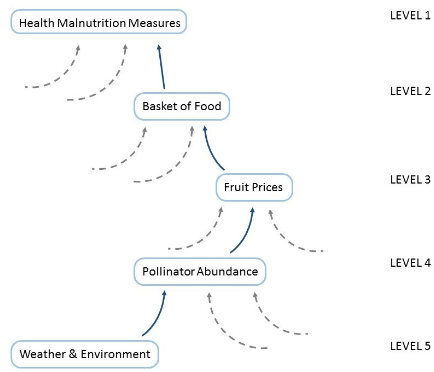

Fig. 2 A visual representation of the levels of inputs and outputs in our Food Security example.

3.2.2 Who can inform these attributes and how?

We then need to *identify experts* who might understand how to *forecast* the future value of each utility attribute vector, albeit in the light of some necessary input information on these features, see the next point. Note that this elicitation will need to be framed within the granularity of the support system as discussed above. These experts will provide a provisional list of further candidates to populate the panels, as described in bullet point 4 in Section 2. Note that sometimes it can be appropriate for a single panel to be responsible for more than one utility attribute vector.

3.2.3 Firming up meaningful inputs and outputs

Representatives of the expert panels are then asked what past or contemporaneous values of the attribute vector or other “Level 2” features (see [40]) they would need to know before being able to calculate the forecast distribution of the utility attribute they take responsibility for - again over the required granularity. The newly identified Level 2 features must then be firmed up and clarified so that they are able to pass the clarity test [17, 18]. If the panel is supported by its own domain specific probabilistic software then this question is often easy to answer. Since such software is usually more fine grained than is needed by the encompassing IDSS, the required attribute forecasts are often simple aggregates - across time and variables - of statistics already calculated by the software. However, in this case our task is often simply to ask the panel to use their software so as to contribute various means and variances conditional on the values of their required inputs. This, for example, is true of many of the economic components used in our Food IDSS. On the other hand it is also often the case that no such probabilistic model is currently available. In this situation the IDSS designers will need to construct a probabilistic module that is able to do this job. Later in this chapter we will describe in some detail how we have performed this task for the pollination services module within the food IDSS. We will see that this usually involves the construction of another qualitative sub-system - often itself either a BN or dynamic BN (DBN). First, new provisional panels need to be identified that can discuss and describe these processes. We will then need to elicit from these discussions a vector of measurement variables that can act as a surrogate for the expert's elicited understanding of the underlying processes. Out of this construction the experts will need to specify inputs it needs to be able to forecast its output. In the pollination example below, one of these concerns different types of weather conditions provided through meteorological forecasting systems, determining how insect pollinator abundance might fluctuate within a given year.

3.2.4 Iterations to provide causal chains

We now take the Level 2 vectors of input measurements needed for the forecasts above and repeat the process substituting these clearly defined vectors for the attribute vectors above. The collections of any vectors needed as input to these models that have not already appeared as either Level 1 or Level 2 vectors we label as "Level 3" vectors. So in the weather example above these new components might involve various measurements of climate change like average earth temperature that will affect weather changes.

We iterate this process, deepening the levels until all input variables have currently known values or are sufficiently remote from the attributes of the process to not have a major impact on the forecasts needed by the system to determine high expected utility scoring policies.

3.2.5 Example

Within our Food IDSS one attribute of the utility function of the the centre is health of the given population as a function of possible malnutrition.

The predictions of this vector of health indicators appear as functions of other "input" variables. One of the components of this input variable is the cost of feeding a household, based on the cost of an appropriately defined basket of food for a household (a Level 2 variable), which passes the clarity test, [17]. To provide an appropriate joint distribution over components like these will be the primary task of our team. In this sense it will constitute one of the attributes of this expert panel's utility.

One of the inputs needed to forecast this price is a specific clarity tested measure of the abundance of particular fruits which will be processed into products in the basket of food above (a Level 3 variable). These in turn will be influenced by pollinator abundance measures (a Level 4 variable). These measures will need models as functions of other Level 5 variables, like weather and environmental factors.

Once this process has been completed the centre will have elicited a collection of random vectors partitioned into depth levels where the level of this depth reflects its distance from the attributes of the decision centre's utility function.

Remark 5. Within this process it is absolutely critical to ensure that the outputs delivered by expert panels at deeper levels of the process *precisely* match the input requirements of the receiving expert panel. Since the donating panels and the receiving panels work in different domains there is a clear danger of confusion at the interface. Unlike in many expert systems, note that in these composite systems the output variables delivered by the expert panels are determined by the needs of the *receiving* panel and are not self-determined. This helps simplify the system, but the needs of the receiving panel can be unfamiliar to the donating panels. Sometimes quite deep discussions are needed between representatives of the adjacent panels before this delivery can be successfully understood and addressed. In extreme cases this might require the panel to develop some interface software, using their domain knowledge in conjunction with a statistical model to transform their standard output to the needs of the composite system.

Remark 6. It can be shown both in a formal sense and also empirically that the distribution of variables furthest away from any attribute tends to have the least effect on the scores of competing candidate policies. So often ignoring uncertainties and using a naïve plug in estimate for these variables, based on for example official statistics and predictions, sample surveys or estimates from an expert who know the field, will be sufficient. An experienced analyst will be able to guide the decision centre when sufficient levels are in place. Typically we would go one level too deep so that deliberations of how things are affecting each other and any associated significant dependences are informed by this extra level.

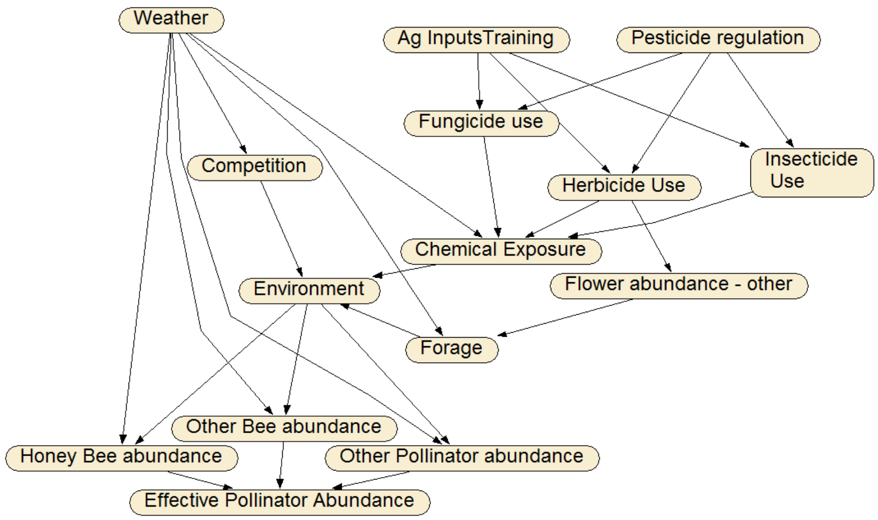

Fig. 3 In this fragment of the network, agricultural inputs training and pesticide regulation operate through several layers of nodes which have more direct effects on pollinator abundance, so neglecting the uncertainty in the estimates for these will have a less deleterious effect on the utility scores that neglecting the uncertainty on the estimates of, for example forage or weather. Produced in Netica, [28].

Notice that the random vectors informing levels of the system described above are often in practice vectors that will need to be indexed by time. So, for example, pollination by bees is affected by weather conditions the previous winter - affecting colony overwintering survival - and also contemperaneous weather conditions affecting bee activity levels. We often need to study and depict these different categories of relationships in the next stage of the elicitation process.

3.3 Listing measurements in a Causal Order

Following the process described above we obtain a hierarchy of well defined random vectors $\{\mathbf{X}_{1,t}, \mathbf{X}_{2,t}, \ldots, \mathbf{X}_{n,t}\}$, where $t = 1, 2, \ldots$ for the prediction of the process in the coming time period. The vectors associated with a given time period are called a *time slice*. Each of these vectors can be associated with a delivered output from one of the panels. In both our running example of the overarching food system and pollinator sub-network, the natural forecast time periods turned out to be a year. Note that the overarching model need not have the same time period as component models. Here, for reasons that will soon become apparent, we index these vectors so that the last vectors in the list corresponded to the attributes of the decision centre's utility function, whilst the earliest random vectors correspond to the random vectors in the deepest levels of the process, $\mathbf{X}_{i,t}$, conventionally called its *parents*, as $Pa(\mathbf{X}_{i,t})$, $i = 1,2,\ldots,n$,. Notice that from the construction and labelling above $Pa(\mathbf{X}_1)$ is the null vector - because by definition $\mathbf{X}_1$ will need no inputs from other components in the system. We call those vectors $\mathbf{X}_i$ for which $Pa(\mathbf{X}_i)$ is the null vector, $i = 1,2,\ldots,n$, *founder vectors*.

We have consciously chosen the elicitation above to be consistent with the principle that it is easiest to think about a vector by seeing this output as an effect whilst inputs can be loosely seen as causes. We are now able to construct pictures that depict processes consistent with this perceived causal structure. For example, obviously it will not be possible to use variables associated with future vectors along with predictive elements from the past and now at the current time. Thus suppose at each time t there are K_l vectors on level l, where $l = 1,2,\ldots,L$ and let

$$\overline{\mathbf{X}}_{lt} = (\mathbf{X}_{1lt}, \mathbf{X}_{2lt}, \ldots, \mathbf{X}_{klt}, \ldots, \mathbf{X}_{Klt})$$

Definition 2. We call a listing *causally compatible* if under the notation above inputs of the component having $\mathbf{X}_{klt}$ as an output must be indexed before $\mathbf{X}_{klt}$.

For simplicity we will henceforth assume here that for our particular IDSS the elicitation admit at least on casually compatible ordering. This will enable us to represent the over arching process as a DBN.

Remark 7. If this condition cannot be met initially then we can often induce it by choosing a finer temporal step. Alternatively, if this transformation does not work then it is sometimes possible to omit some of the entries in the parent set corresponding to elicited contested or weak dependences. If neither of these reconstructions is possible then we will need to represent the problem by a reciprocal graph [19]. The formal methods we describe below then still apply. However, the outputs of the system are then less transparent and the calculations we can make directly from the system are far more costly and time consuming to make.

Through the construction above we can allow that contemporaneous vectors $\mathbf{X}_{klt}$ on the same level to depend on one another. However, the causal compatibility condition above ensures that inputs of the component having this as an output must be indexed before $\mathbf{X}_{klt}$. With this constraint we can now choose a listing of variables time slice by time slice as follows:

Time	1	1	$\ldots$	1	2	2	$\ldots$	2	$\cdots$	T	T	$\ldots$ T
Depth	L	$L-1$	$\ldots$	1	L	$L-1$	$\ldots$	1	$\ldots$	L	$L-1$	1
Vector	$\overline{\mathbf{X}}_{L1}$	$\overline{\mathbf{X}}_{(L-1)1}$		$\overline{\mathbf{X}}_{11}$	$\overline{\mathbf{X}}_{L2}$	$\overline{\mathbf{X}}_{(L-1)2}$		$\overline{\mathbf{X}}_{12}$		$\overline{\mathbf{X}}_{LT}$	$\overline{\mathbf{X}}_{(L-1)T}$	$\overline{\mathbf{X}}_{1T}$

Since through our elicitation process we preclude the possibility that higher time indexed outputs can serve as inputs to lower time indexed ones, by concatenating the vectors $\overline{\mathbf{X}}_{lt}$ in the order above we obtain an ordering of the vectors that will describe a causally compatible listing of vectors.

3.4 Bayesian Networks and Dynamic Bayesian Networks

The IDSS model works for many different frameworks depending on the underlying purpose of the DSS. For our food security application we have chosen to use DBNs and we shall therefore define this specific type of graphical model in this section.

3.4.1 Defining a graph

Once this process is completed it is straightforward to express these elicited dependence statements graphically.

Definition 3. A Bayesian Network of a causally compatible listing of random vectors has as its vertices the set of these component vectors. A directed edge exists from $\mathbf{X}_{k'lt}$ into $\mathbf{X}_{klt}$ if and only if $\mathbf{X}_{k'lt}$ is a component of $Pa(\mathbf{X}_{klt})$, with $Pa(\mathbf{X}_{klt})$ as defined above.

In [40] we proved that this is indeed a Bayesian Network. In fact it is also a Causal Bayesian Network in the sense of [31] and can be used as the basis of a DBN. So in particular, not only does this represent the genuine beliefs of the expert panels about how one component of the system affects another, it is not merely a representation but has a formal interpretation. This means that the directed graph itself - representing the composite of dependences expressed in terms of *individual local judgements* - can be interrogated for its plausibility of the logically implied global picture of interdependences as a whole! So once the first parse of elicitations takes place we can examine, through for example the use of the d-separation theorem ([31, 40]), whether the composite structure representation really can stand up to scrutiny. The way that such an interrogation might proceed is discussed and illustrated in [40].

Finally, because of the formal compatibility of the system, after a sequence of interrogation steps pick out a structure which to all responsible parties seems plausible we can immediately use this graph as a framework for constructing a completely specified stochastic process. This is done by eliciting, for each vector $\mathbf{X}_{kLt}$ in the system, a probability distribution of this vector conditional on each configuration of its parents. Explicitly the joint density of the whole process can be defined by

$$p(\mathbf{x}) = \prod_{k,l,t} p_{k,l,t}(\mathbf{x}_{k,l,t}|Pa(\mathbf{x}_{k,l,t}))$$

where $p_{k,l,t}(\mathbf{x}_{k,l,t}|Pa(\mathbf{x}_{k,l,t}))$ are simply the conditional density or - in the discrete case - condition probability mass function or table of $\mathbf{X}_{k,l,t}|Pa(\mathbf{X}_{k,l,t})$. Each of these components in the formula may already be available from existing probability models specific to a given domain. In reality once the system has been specified we will inevitably find gaps where no such formal analysis has taken place. So to fill the gaps, to obtain quantitative measures of the required expected utilities, we need to elicit such distributions directly: see pollination example below. Often these distributions will actually be margins of other Bayesian networks.

Remark 8. When experts design their own systems, sometimes the *internal structure* of one component can share variables with the internal structure of another. So, for example, flooding could disrupt both the production of food and its distribution and yet these might be forecast using different components. In such cases it is important that such shared information is provided externally though a different module, possibly overseen by a different expert panel. In our example this might be a meteorological panel.

3.4.2 Feasible Graphical Models and Simplifying Structures

The construction described above, whilst formally powerful would also lead us to build directed graphs which, in the contexts we are describing above might have hundreds or even tens of thousands of vertices. Obviously, as pictures, such graphs are very difficult to read or synthesise! Furthermore, the elicitation of the various conditional probability tables or distributions needed for the system would be prohibitive. However, there are now various techniques available to reduce the number of specifications and to make the depiction of the underlying processes more accessible.

There are two established collections of assumptions which restrict the underlying BNs to take a particular form and so vastly reduce the necessary inputs to the system which make even very complex models amenable to a parsimonious analysis. There are many of these frameworks and the appropriate choice will obviously depend on the underlying purpose of a given IDSS. The two families of model are; the multiregression dynamic model (MDM) [35] whose use in conjunction with these complex dynamic systems is discussed in, for example [23, 41, 42] and the two time slice dynamic Bayesian network (2TSDBN), see [21]. Accommodating various additional assumptions about the process both reduce the picture to a graph over one or two time slices. Since the 2TSDBN is less technical and so easy to describe and is used in our example we will detail only this particular construction in this chapter.

The 2TSDBN simply elicits the DBN up to the variables in the second time slice. It then makes an additional Markov assumption: it assumes that all the parent components can be written as

$$Pa(\mathbf{x}_{k,l,t}) = \left\{\mathbf{x}_{k',l',t'} : k' \in K', l' \in L', t' = t \text{ or } t-1\right\}.$$

The key point here is that we assume once dependencies between contemporaneous values of measurements and the most recent past values are known, then the more distant past provides no further useful information. Furthermore, we assume that the process is time homogeneous after the first time point. Suppose the time step is a year. Then under these assumptions determining the conditional distributions associated with processes with what will happen next year and how this is affected by what is happening this year, the process can be fully specified see Figure 4 and Figure 5. In particular, to depict the process we simply need to specify a graph on two time slices: containing vertices associated with measurements for this year and measurements for next and the dependencies between variables in the two time-slices. We illustrate this in the example below.

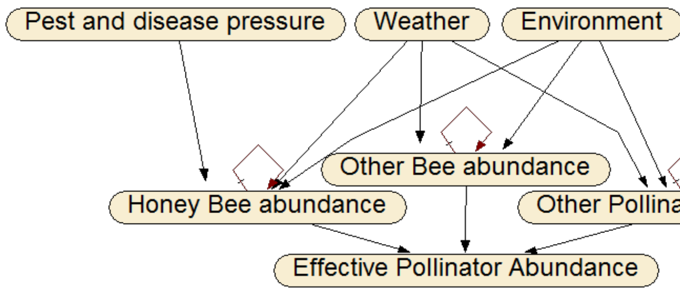

Fig. 4 We model the fact that pollinator abundance in the current time is influenced by pollinator abundance in the previous season, through the numbers entering overwintering. This is shown as a self-loop here, but this violates the DAG requirements. Produced in Netica, [28].

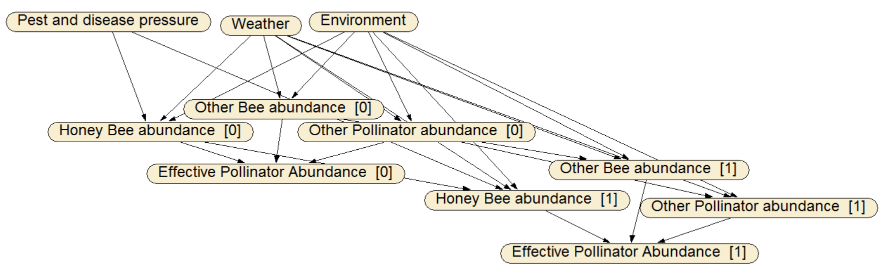

Fig. 5 Here the 2 time-slice DBN has been expanded in time. Note that, as well as weather and environment effects, at time [1] honey bee abundance is affected by honey bee abundance at time [0] and the same is true for the other insect pollinator types. Produced in Netica, [28].

Each node in our schematic, shown in full detail in Figure 1, represents a list of random variables. An arrow is drawn between nodes when *at least one* variable is causally connected to at least one other variable in another node, or if there is a temporal relationship present. For specific classes of BNs such as Object Orientated Bayesian Networks (OOBNs) we usually begin with the probability distributions and then group similar objects together to create the overarching class nodes, moving from left to right in Figure 6. Note that the schematic for Figure 6 can be formally interpreted as a BN provided that we understand the schematic as representing a BN *conditional* on the relevant past variables from the time $t-1$ slice. Such conditional BNs have recently proved a useful modelling tool in other contexts, see for example [29]. However, in our application we are trying to construct our BN using literature and experts, so we first use common language to derive a general schematic and then more formally break these class nodes into more specific and detailed probability distributions. Working from a coarse to fine level in this way is a much more natural process for applications such as ours and is equivalent to moving from right to left in Figure 6.

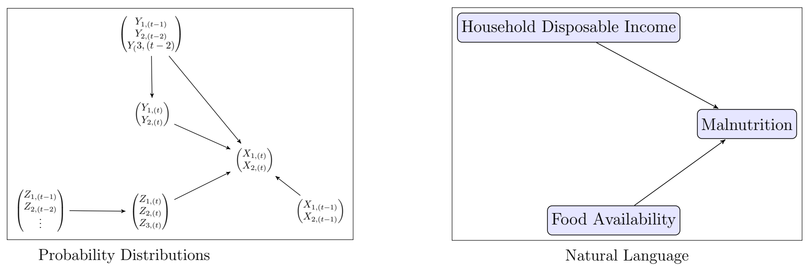

Fig. 6 Process of moving from natural language, Figure 6(a), to probability distributions, Figure 6(b), or vice versa.

Of course the formally elicited graph is still there and can be used for formal deductions and construction. The simplified summary graph is nevertheless useful for depicting some important features of the elicited dependence structure and feeding this back to the panels for discussion and verification.

Remark 9. When studying and verifying these elicited pictures of the elicited process of the process it is often useful to compare a structure with various mind maps and other schematic depictions informed by deep reflection by experts in the field. Although such pictures of dependences rarely have a formal interpretation, they are often the result of deep reflection and provide supporting narratives which might produce compelling reasons for adopting modified dependencies. Furthermore, they are useful for helping populate provisional initial lists of vectors on which the process needs to be built.

Remark 10. Before the elicitation starts it is always necessary to do some preparatory work. With the help of various friendly domain experts, the analyst will need to trawl any relevant literature and check which hypotheses found there might still be current.

In this way, the elicitation of the qualitative structure of the determinants of household food security in the UK began with the decision-makers, local government, determining the attributes of their utility function.

Beginning with the schematic structure of the UK food system presented in [6], a trawl of relevant literature was used to construct a plausible model of the qualitative structure, starting with the Level 1 elements which influence directly the attributes of the utility defined by the decision-makers. This qualitative structure was then iterated in consultation with food poverty domain experts from nutrition, politics, sociology, crop science and the local authority decision-makers in person and using the reports and other resources they recommended. A final version was produced which represents the consensus of these experts.

The qualitative structure for the UK food poverty model was what we thought might be a reasonable summary after an intensive review of the relevant literature. It was presented to various food domain experts from sociology, nutrition, crop science and political backgrounds, individually and in groups, who confirmed the veracity of qualitative structure or suggested amendments to improve the superstructure of the parts of the system they were expert in.

We stated that is is often necessary, especially in the context of food poverty modelling, that the designers of the system will have to enable the panel to build a bespoke probabilistic model as one of the components. Perhaps the most straightforward way of doing this is to use DBN technologies with their supporting softwares to define this, especially when the probabilities expressed in the processes represent expert judgements. From a coding perspective we then have a DBN "object" that represents a node and its dependency structures.

Over the last two years this element of our research programme has been so important we will spend the next section describing one such elicitation in some detail. The process for the elicitation of the sub-component is more focussed and fine-grained, but otherwise identical to the elicitation methodology for the overarching system. So in this way, since both structures we choose to define here are dynamic forms of the BN, we can kill two birds with one stone.

So we next describe the process of eliciting the structure of the process leading to pollination, by a panel who were expert in the process of pollination as it applied to crop yield. This output would then inform any inputs associated with both the availability and price of certain items in a basket of food available to a person with stretched means and hence their health, educational attainment and social discontent. The analysis gave some of the probabilistic judgements we needed with a suite of BNs that were needed in relationships between the broad category labels such as Weather & Environment and Farming, displayed in Figure 1.

4 Bayes Nets for a component model: A Case study

Embedded in the crop production element of the UK food security model is a need for pollination services. A large proportion of important food crops are insect-pollinated and the current concern about falling numbers of pollinators and the impact on food production means this is an important element to model. There is also a need for decision support for those charged with ensuring the implementation and ongoing development of the UK's National Pollinator Strategy. However, there is considerable uncertainty and a dearth of evidence for some key parts of the DBN representing the pollination system, so for this we conducted a structured expert elicitation using the IDEA protocol: described in a Chapter 4 of this book.

4.0.3 Development of the Bayesian network structure

The UK pollinator strategy recently published by DEFRA provided the backbone for the development of the implicit utility function (pollinator abundance) and the relevant expert panels to provide evidence on factors influencing pollinator abundance. The abundance of pollinators can be subdivided into three categories; abundance of managed bees, abundance of other bees and abundance of other pollinators. Having established that pollinator abundance with regards to pollination of UK crops is the target of our model, the variables affecting this directly and indirectly were then identified in an iterative process.

The first draft of the BN was produced by academics, see Figure 7, including one of the authors of this chapter, contributing respectively, expertise on BNs, honey bee disease dynamics and species distribution models. Using their background knowledge the first sketch was drawn, challenged and re-drawn until the underlying probability statements implied by the structure seemed plausible. The variables identified at this stage were the availability of forage, suitable nesting sites and the prevalence of disease. These in turn were influenced by the weather, pesticide use, competition and land use including crop type distribution. More detail and documented research is available on the diseases, parasites and predators affecting honey bees than solitary bees and other pollinators, and this is reflected in the first draft of the BN, Figure 7.

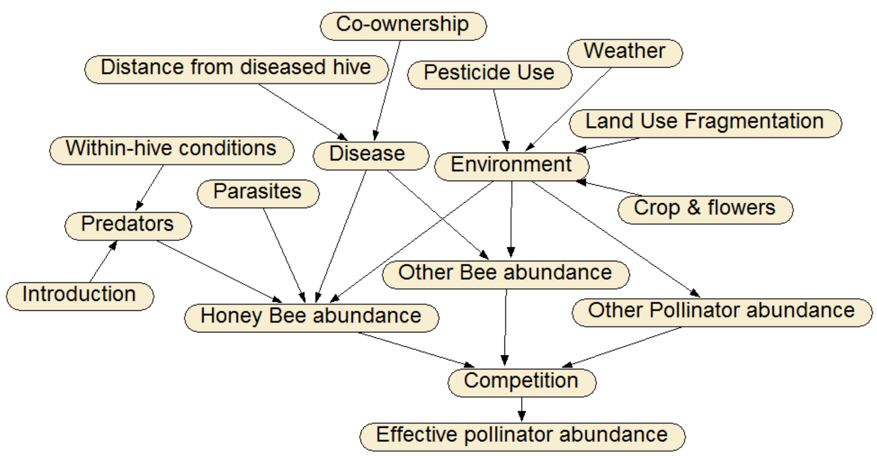

Fig. 7 2014: first draft of DBN with academics. Produced in Netica, [28].

The second step was to search the academic literature on pollinating insects and incorporate the new insights gained into the next iteration of the BN. It became clear that weather conditions affect the dynamics of the system on many more levels than previously thought, for example affecting the prevalence of parasites of honey bees and the availability of forage. Through its effects on weeds, fungi and insects that attack food crops, weather also influences the specific pesticide product employed, which in turn may affect pollinating insects. This was incorporated into the second draft, Figure 8.

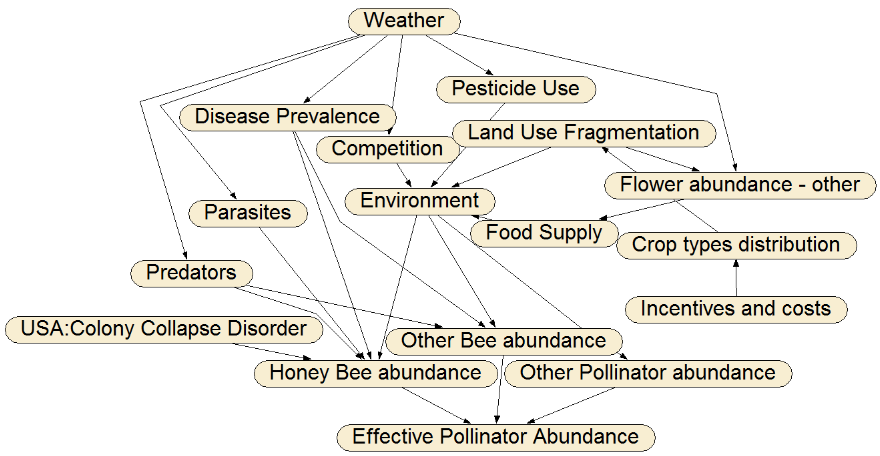

Fig. 8 The second draft incorporated insights from a detailed search of the academic literature. Produced in Netica, [28].

The third stage was to conduct a series of interviews with pollinator experts, undertaken by one of the authors of this chapter, ensuring we included as many different types as possible: beekeepers, government agency experts, Queen breeders, academics, honey producers, wild bee experts and government researchers. The next iteration of the qualitative structure included the direct effect of weather on the insects themselves - bees are prompted to forage by daylight and temperature and colonies can fail to survive if a winter is too long or too cold and therefore weather also affects numbers directly. The experts also informed us that the BN should include human factors; in the case of honey bees, the competence of the bee-keeper and for the environment, some measure of its management. As many of these interviews as possible were carried out face-to-face and the experts, having had the network explained to them, indicated changes that needed to be made or annotated the network themselves as in Figure 9. For those who were available only by telephone, a list of questions were constructed to guide the conversation, but they were also allowed to comment freely from their perspective. Further resources recommended by the experts, such as government reports, policies, research articles and other experts, were followed up and the structure adjusted accordingly. Happily, there were no mutually exclusive opinions expressed, so the network structure expresses the consensus of the domain experts. These interviewees also helped to populate the decision space with the candidate policies which could be enacted with an assessment of the likely efficacy under given circumstances, although some of the interviewees were from overseas and so some of the strategies were not appropriate to the UK due to the varying bee species and ecological systems. These interviews were also very useful in helping to refine the definition of each of the variables which informed us how they could be measured. Finally, a DBN expert with extensive experience in using DBNs for decision support in an ecological domain suggested that the disease / treatment pairs be explicitly included, to evaluate fine-grained specific interventions for example the inclusion of Antibiotics, Miticides and Pesticides nodes, Figure 10.

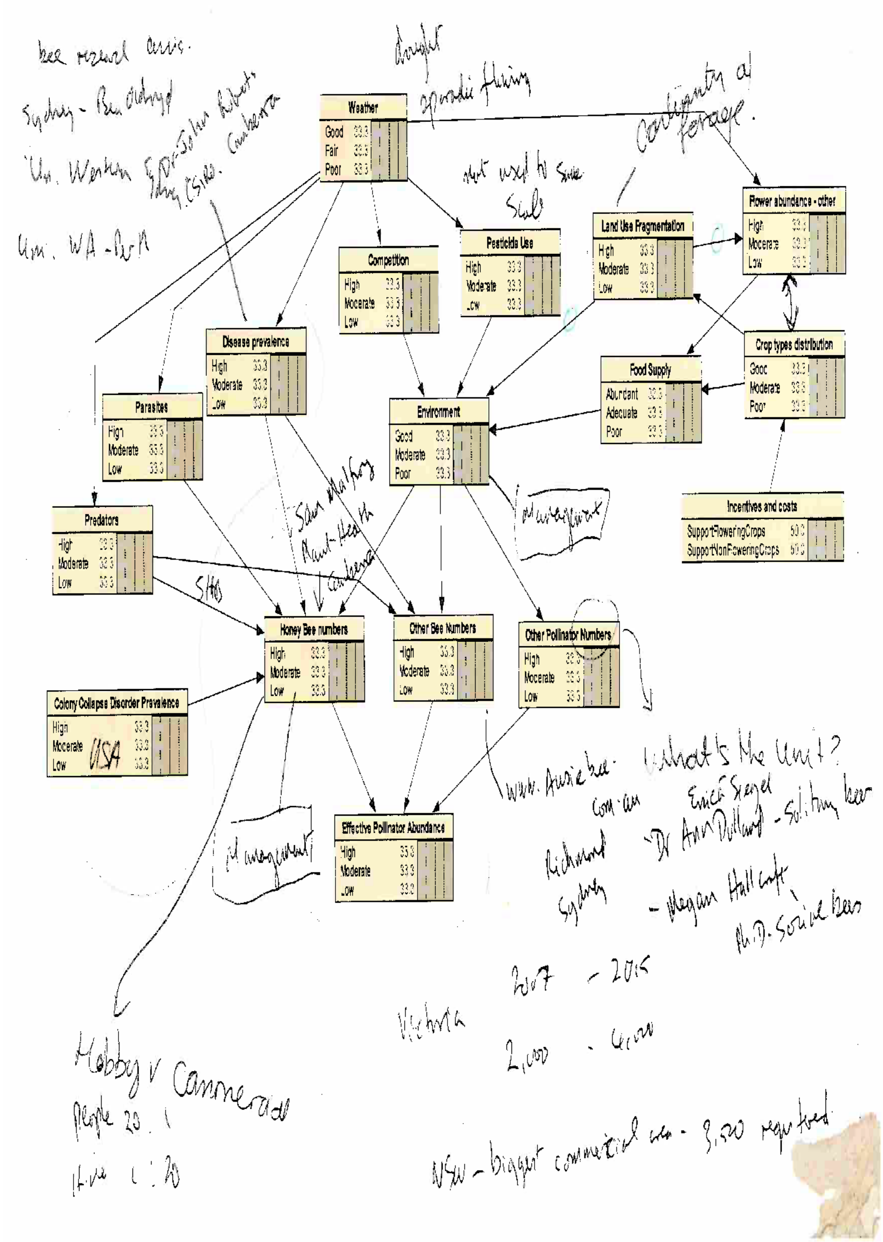

Fig. 9 Some experts annotated the network themselves

The final addition to our network was the time dependencies (seen in Figure Figure 10 as red loops or lines) - all variables influence each other within the same season, but the abundance of pollinators entering the over-winter period directly affects their abundance in the next season and similarly bee-keeper training affects management in each subsequent season as well as the current one, so these are represented by time-delay links in the DBN.

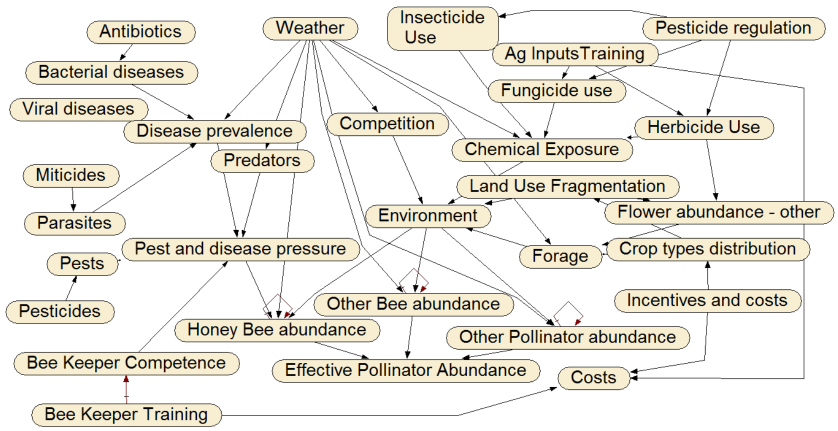

Fig. 10 Fully detailed DBN, including the time dependencies. Produced in Netica, [28].

This completes the structural elicitation phase of this component.

4.1 Eliciting CPTs

Having established the qualitative framework and checked the implied conditional probabilities make sense to domain experts, the next task was to populate the conditional probability distributions of the variables at each node. The academic literature is able to provide some estimates of these, particularity the marginal probabilities, but the conditional probabilities were not so readily available. The IDSS theory developed in [41] shows that we can legitimately and coherently admit expert judgement as evidence in an IDSS alongside experimental and observational studies. We therefore turned to expert judgement to populate the conditional probability distributions for the parts of the system with least evidence and most uncertainty.

Experts in pollinators were gathered for a structured elicitation exercise, using the IDEA protocol (see Chapter 4), to provide probability distributions for the effects of weather, environment and disease on abundance of honey bees, other bees and other insect pollinators, such as hover flies. The fragment of the Bayesian network for which quantities were elicited is shown in Figure 11.

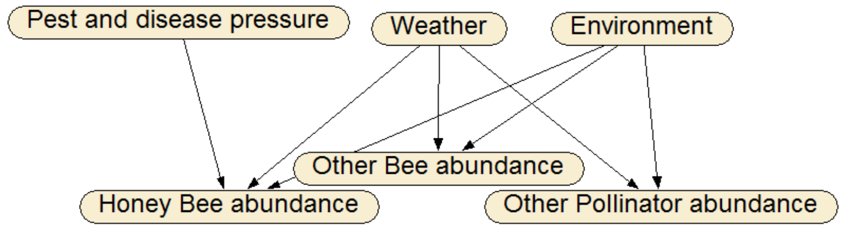

Fig. 11 Expert elicitations was used to derive conditional probability distributions for the parts of the system with least evidence and most uncertainty. Produced in Netica, [28].

The experts were identified with the assistance of the government agency responsible for pollination in the UK including advice to identify experts from within the insect pollinator initiative.

Eleven experts participated in the workshop, their expertise covering honey bees, plant-pollinator relationships, bees and farming, hoverflies, pollinator viruses, wild pollinators and mathematical epidemiology, including epidemiology of honey bee diseases and pests. A list of resources was circulated as background reading, to make the evidence available equally to all experts, and experts also forwarded suggestions for further resources which were circulated too.

One expert also attended for the previous day to lend domain knowledge to the definition of the variables and refinement of the questions of interest to ensure that they made sense to the domain experts, and were precisely and unambigously worded. The workshop started with a presentation of the problem and presentation of the IDEA protocol. The question of biases was explained and the importance of answering the questions in the order they are posed, in order to avoid anchoring, was also discussed. The definitions of the variables were scrutinised and there were some changes made at the request of the assembled experts. The most significant change was to the definition of a supportive environment which was changed from 'at least 1 hectare of open flowers' to 'at least 15% proportion of semi-natural land', a measure more routinely used in domain literature. The experts then gave their individual first-round subjective estimates of the lowest plausible, highest plausible and best estimate of probabilities of interest, 16 combinations in all.

The subjective estimates were collected and data entered in anonymised form to produce graphs for each question showing each experts estimates and upper and lower bounds and the group mean for each question side-by-side on a graph. A facilitated discussion of results followed and each expert had the opportunity to contribute and give reasons for their estimates. Discussion was encouraged especially when different opinions were depicted in the graphs, but also when agreement was observed. During the discussions experts were encouraged to note down anything that they considered to be significant new information or compelling argument, especially if it had altered their thinking about a question.

After the close of the discussions, the experts gave a second set of individual, subjective estimates for the same questions, having their original estimates returned to them for comparison. Finally, an on-line protocol was discussed for running calibration questions, where the experts would be given several days to fill in the first round estimates, have an on-line facilitated discussion and then give a second round of estimates for the calibration questions. The experts indicated that they understood the value of this element of the elicitation to the final results and all agreed to participate. Following this, the experts' judgements were aggregated.

In this way, the first versions of the DBN of pollinators has been built. Of course, this may need further refinement once the functionality of the composite system is calibrated for plausibility - see the next section. However, once completed we have a system which takes inputs, for example measures of the use of antibiotics, weather variable and pesticide regulation, taken as known or provided by other panels and it provides an output for other parts of the model to use here.

The probabilistic output of this pollination sub-module provides the CPTs needed for part of the components of the overarching food system, the "Farming" module, represented in Figure 1 above. Once all the components of the "Farming" model have been populated with probabilities in analogous ways, the "Farming" module will have a fully specified set of CPTs conditional on its input variables, here listed under the broad heading of Economy and Weather & Environment.

Finally, this process needs to be repeated for each of the vertices in Figure 1. We have then elicited a probability model of features in the process needed to score various policy options. The population of this model is obviously a massive task. However, our small team has made significant strides in producing this IDSS which we plan to deliver in prototype form in two years time.

5 Quality Control of Integrating systems, Diagnostics & Robustness

In any complex model it is absolutely critical that once built, the integrating system produces credible composite forecasts of the processes of interest. This is true of each subcomponent, component and of the whole IDSS itself. Even when the overarching system is requisite to everybody and individually each component appears plausible, at least from a theoretical point of view it is quite possible that vital components or dependencies in the structure have been missed.

However, because the integrating system has been modelled so that it is probabilistically coherent, standard methods for checking the robustness and diagnostically checking any probabilistic system can be devised to address this domain. Here we simply have a massive system, but the underlying suites of methodologies are the same.

There are various methods we can use here to check out whether the integrating system is fit for purpose and if not to modify it so that it is. These are often called *diagnostic checks*. Here we illustrate quality checks that we will perform on the Food IDSS, once fully populated, which we inherited from other modelling activities.

Perhaps the most important consistency checks are ones that ensure that the IDSS works *predictively* well. This is because it is these distributions which will determine the way the centre scores various options open to it. The easiest element of this is to check one-step ahead forecasts with observations that are actually seen. This can be performed by retrospectively using the methods on a suitable number of past time points, using informative forecasts available at the time. These can be constructed retrospectively and the system run against this systematic, for example over or under estimation of attribute vectors can then be identified.

We can also compare future predictions from our model to predictions that the experts forecast. This would enable us to use any new information in the formulation of the model and would also highlight any discrepancies which occur, for example our model may predict certain outcomes which the expert knows will never occur. This can then be fed back into the system, helping to improve future forecasts.

Our IDSS systems are usually intended for use by policymakers within governmental agencies or industrial companies who may not include expert statisticians. For example, our food IDSS is being designed for, and in collaboration with, local councils within the UK. It is therefore imperative that we build into the software on-line diagnostics and robustness alerts, so that red flags instantly appear if there is a potential error in the information entered or if an outcome is predicted which is infeasible. This will enable the user of the system to note any discrepancies and account for any additional uncertainty they may obtain due to these causes.

Diagnostic checks for BNs have been studied over the years, and we shall briefly discuss three in particular which were introduced by [10] as a way of quantifying the difference between the specified prior and the data. We shall briefly discuss three of their diagnostics here: a parent-child monitor, a node monitor and a global monitor

1. The *Parent-Child Monitor* examines whether the priors in each of the CPTs of a given node are accurate given the parent nodes.
2. The *Node Monitor* quantifies the performance of predictions on a specific node, given all available evidence.
3. The *Global Monitor* determines the overall performance of the graphical model.

These diagnostic checks all aim to quantify how well the model is performing after the data is observed. This measurement of performance is done using formula such as scoring rules.

Any real probabilistic system has a degree of misspecification in it, either within its assumed structure or the specification of the probabilities within it. Of course, one way of addressing this is to perform a one at a time sensitivity analysis. However, when examining the robustness of the types of massive system we consider here, limited information about robustness can be gleaned in this way. Rather it is better to consider the robustness of the outputs to perturbations of the whole system. One spin-off of a formal analysis of this type is that we can discover *before* the BN is fully elicited that the precision of probability specifications in parts of the system can have little impact on the process *however* we fill in the CPTs. In this case little effort needs to be expended on these elements of the process.

We note that if we are deciding between two models, Bayes Factor methods can be utilised very simply to appreciate how much gain can be obtained in explanatory power from using a more complicated model rather than a simpler one. In our context here, all things being equal, we would advise the choice of a simpler model over a complex one if the gain in using the complex one is marginal. These techniques are widespread for standard graphical models, see [10, 21, 40], and are currently being examined in this specific context by two of these authors. Early results suggest that relative scores are most affected by misspecification of input distributions and structural features close to the attribute vector of the graph in the BN specification of the process.

6 Conclusions

We have demonstrated that building a fully probabilistic IDSS is feasible and been able to communicate some of our experiences in engaging in this activity. We hope that others will follow our lead and start to produce similar tools to address other large scale decision analyses. We believe that surmounting the challenges of implementing such large scale tools will be increasingly important in the coming years.

References

1. Becher, M. A., Osborne, J. L., Thorbek, P., Kennedy, P. J. and Grimm, V., 2013. Towards a systems approach for understanding honey bee decline: a stocktaking and synthesis of existing models. *The Journal of Applied Ecology, 50*(4), pp. 868-880.
2. Bishop, J., Jones, H.E., Lukac, M. and Potts, S.G., 2016. Insect pollination reduces yield loss following heat stress in faba bean (Vicia faba L.). *Agriculture, ecosystems & environment*, 220, pp.89-96.
3. Blaauw, B.R. and Isaacs, R., 2014. Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *Journal of Applied Ecology, 51*(4), pp.890-898.
4. Breeze, T.D., Roberts, S.P., Potts, S.G. and Potts, S.G., 2012. The Decline of Englands Bees: Policy review and recommendations. *University of Reading*.
5. Caminada, G., French, S., Politis, K. and Smith, J.Q., 1999. *Uncertainty in RODOS*. Doc. RODOS(B) RP(94) 05.
6. Collier, R. A., 2009. Identify reasons why food security may be an issue requiring specific attention. DEFRA Research Project Final Report.
7. Chauzat, P., Laurent, M., Rivire, M.P., Saugeon, C., Hendrikx, P., Ribire-Chabert, M. and pathology Unit, H., 2014. A Pan-European Epidemiological Study on Honey Bee Colony Losses 20122013. *European Union Reference Laboratory for Honeybee Health*.
8. Datta, S., Bull, J.C., Budge, G.E. and Keeling, M.J., 2013. Modelling the spread of American foulbrood in honeybees. *Journal of The Royal Society Interface, 10*(88).
9. Dawid, A.P., 2001. Separoids: A mathematical framework for conditional independence and irrelevance. Annals of Mathematics and Artificial Intelligence, 32(1-4), pp.335-372. Vancouver
10. Dawid, A.P., Cowell, R.G., Lauritzen, S.L. and Spiegelhalter, D.J., 1999. *Probabilistic networks and expert systems*. Springer-Verlag, New York.
11. Edwards, W., Miles, R.F. and Von Winterfeldt, D., 2005. *Advances in decision analysis.* Cambridge University Press.
12. EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment, *European Food Safety Authority (EFSA) Journal, 12*(6).

13. French, S. and Maule, J. and Papamichail, K. N., 2009. *Decision Behaviour, Analysis and Support*. Cambridge University Press.
14. French, S. and Smith, J., 2016. Decision Analytic Framework for a Decision Support System for Nuclear Emergency Management. In *UK Success Stories in Industrial Mathematics (pp. 163-169)*. Springer International Publishing.
15. Gordon, R., BresolinSchott, N. and East, I.J., 2014. Nomadic beekeeper movements create the potential for widespread disease in the honeybee industry. *Australian veterinary journal, 92*(8), pp.283-290.
16. Hafi, A., Millist, N., Morey, K., Caley, P. and Buetre, B., 2012. *A benefit-cost framework for responding to an incursion of Varroa destructor*. ABARES.
17. Howard, R.A., 1988. Decision analysis: practice and promise. *Management science, 34*(6), pp.679-695.
18. Howard, R.A., 1990. From influence to relevance to knowledge. *Influence diagrams, belief nets and decision analysis*, Oliver , R.M. and Smith J.Q. (Eds), Wiley 3-23, pp.3-23.
19. Koster, J.T., 1996. Markov properties of non-recursive causal models. *The Annals of Statistics, 24*(5), pp.2148-2177.
20. Khoury, D. S., Barron, A. B. and Myerscough, M. R., 2013. Modelling food and population dynamics in honey bee colonies. *PloS one, 8*(5), p.e59084.
21. Korb, K. B. and Nicholson, A. E., 2011. *Bayesian artificial intelligence*. CRC press.
22. Lagi, M., Bertrand, K.Z. and Bar-Yam, Y., 2011. The Food Crises and Political Instability in North Africa and the Middle East. arXiv preprint:1108.2455.
23. Leonelli, M. and Smith, J.Q., 2015. Bayesian decision support for complex systems with many distributed experts. *Annals of Operations Research, 235*(1), pp.517-542.
24. Leonelli, M. and Smith, J.Q., 2013, April. Using graphical models and multi-attribute utility theory for probabilistic uncertainty handling in large systems, with application to the nuclear emergency management. In *Data Engineering Workshops (ICDEW), 2013 IEEE 29th International Conference* on (pp. 181-192). IEEE.
25. Leonelli, M. and Smith, J.Q., 2013. Dynamic uncertainty handling for coherent decision making in nuclear emergency response. In *Proceedings of the winter meeting of the ANS*.
26. Lever, J. J., Nes, E. H., Scheffer, M. and Bascompte, J., 2014. The sudden collapse of pollinator communities. *Ecology letters, 17*(3), pp.350-359.
27. Lonsdorf, E., Kremen, C., Ricketts, T., Winfree, R., Williams, N. and Greenleaf, S., 2009. Modelling pollination services across agricultural landscapes. *Annals of botany 103*, pp. 1589-1600.
28. Norsys (1994-2016). Netica. Norsys.
29. Oates C. J., Smith J. Q. and Mukherjee S., 2016. Identification of Conditional DAG Models. *Journal of Machine Learning Research*. (to appear)
30. Pearl, J., 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, San Francisco.
31. Pearl, J., 2000. *Causality: models, reasoning and inference*. Cambridge University Press.
32. Phillips, L.D., 1984. A theory of requisite decision models, *Acta Pschologia* 56, pp. 29-48.
33. Potts, S. G., Roberts, S. P., Dean, R., Marris, G., Brown, M. A., Jones, R., Neumann, P. and Settele, J., 2009. Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research, 49*(1), pp.15-22.
34. Puch, R.O. and Smith, J.Q., 2002. *FINDS:A training Package to Assess Forensic Fibre Evidence*, Advances in Artificial Intelligence , Eds C.A.C.Coella, A.de Albornoz, L.E. Sucar, and O.S. Battistutti , Springer, pp. 420-429.
35. Queen, C.M., and Smith, J.Q., 1993. Multi-regression dynamic models. *J.R. Statist. Soc. B*, Vol.55, No.4, 849-870.
36. Rader, R., Bartomeus, I., Garibaldi, L.A., Garratt, M.P.D., Howlett, B.G., Winfree, R., Cunningham, S.A., Mayfield, M.M., Arthur, A.D., Andersson, G.K., Bommarco, R., *et. al* 2016. Non-bee insects are important contributors to global crop pollination. *Proceedings of the National Academy of Sciences, 113*(1), pp.146-151.
37. Santamara, S., Galeano, J., Pastor, J.M. and Mndez, M., 2015. Removing interactions, rather than species, casts doubt on the high robustness of pollination networks. *Oikos*.

38. Smith, J.Q. and Daneshkhah, A., 2010. On the robustness of Bayesian networks to learning from non-conjugate sampling. *International Journal of Approximate Reasoning, 51*(5), pp.558-572.
39. Smith, J.Q. and Rigat, F., 2012. Isoseparation and robustness in finite parameter Bayesian inference. *Annals of the Institute of Statistical Mathematics 64*(3), pp. 495-519.
40. Smith, J.Q., 2010. *Bayesian decision analysis: principles and practice*. Cambridge University Press.
41. Smith, J.Q., Barons, M.J. and Leonelli, M., 2015. Coherent inference for integrating decision support systems, arXiv preprint:1507.07394.
42. Smith, J. Q. and Barons, M. J. and Leonelli, M., 2015. Decision focused inference on networked probabilistic systems: with applications to food security, *Proceedings of the Joint Statistical Meeting*, pp. 3220-3233.
43. DESA, U., 2015. *World population prospects: the 2012 revision, key findings and advance tables*. Working paper no. ESA/P/WP. 227. New York: United Nations Department of Economic and Social Affairs, Population Division.
44. Vanbergen, A.J., Heard, M.S., Breeze, T., Potts, S.G. and Hanley, N., 2014. Status and value of pollinators and pollination services: a report to the department for environment, food and rural affairs (defra).
45. Xia, A., Huggins, R.M., Barons, M.J. and Guillot, L., 2016. A marked renewal process model for the size of a honey bee colony. arXiv preprint:1604.00051.