⋆

Models in Systems Medicine

⋆

Jon Williamson

Draft of February 2, 2017

Abstract

Systems medicine is a promising new paradigm for discovering associations, causal relationships and mechanisms in medicine. But it faces some tough challenges that arise from the use of big data: in particular, the problem of how to integrate evidence and the problem of how to structure the development of models. I argue that objective Bayesian models offer one way of tackling the evidence integration problem. I also offer a general methodology for structuring the development of models, within which the objective Bayesian approach fits rather naturally.

Systems medicine applies systems approaches, analogous to those used in systems biology, with the aim of improving medical treatment and progressing medical science. These approaches are often described as 'data-intensive' or 'data-driven' because they attempt to draw inferences from a variety of large datasets. This paper explores two problems that face systems medicine. First, there is the problem of diversity of evidence: in addition to large amounts of data ('big data'), the available evidence tends also to be very heterogeneous, and the question arises as to how the whole range of evidence can be integrated in a coherent manner, to enable reliable inferences. The second problem is that of diversity of models: systems medicine employs different models for different purposes, and it is often far from clear as to how these models relate to one another. Can anything be done to shed light on the relationships between models?

This paper develops a normative response to these problems. It puts forward an approach based on Bayesian epistemology for integrating multiple datasets. It then puts forward a way to integrate evidence of mechanisms, which can often be qualitative, into the resulting quantitative models. (This approach can be thought of as a contribution to the EBM+ programme, which seeks ways of integrating evidence of mechanisms with evidence of associations in order to lead to better outcomes in medicine—see ebmplus.org.) The paper goes on to suggest that Bayesian networks can provide a unified modelling formalism. (This conclusion, if not the detail of the approach, is in line with that of Landes et al. (2017), who present a Bayesian network modelling framework for inference in pharmacology.) There is no claim that the framework developed here is the only way to tackle the foundational problems that face systems medicine, but it is hoped that the present attempt will encourage others to tackle these problems.

The paper is structured as follows. §1 introduces systems medicine and notes that its appeal to a wide variety of data makes it a promising new paradigm for medical research. However, progress in systems medicine has not been as rapid as some have anticipated. In §2 it is suggested that this slow progress might be explained by the enormity of the challenges faced by systems medicine. Two challenges stand out as particularly pressing: how should the massive amount of evidence in systems medicine be integrated? how should one go about modelling in systems medicine? In §3 I classify models in systems medicine as being of four kinds: quantitative models of association; quantitative causal models; qualitative mechanistic models; and quantitative mechanistic models. In §4 I show how objective Bayesian epistemology can be applied to data integration and how an objective Bayesian net can be used as an association model. In §5 I then sketch a principled way of generating a causal model, and of structuring the development of models in systems medicine in general.

§1 Systems Medicine

Systems medicine and systems biology. Systems medicine is an approach to medicine that has emerged only in the last few years. Systems medicine is closely related to systems biology, which studies biological systems holistically. Typically, systems of molecules and their interactions within the cell are the primary objects of study of systems biology, and its main aim is to discover new biological mechanisms (see Boogerd et al., 2007, e.g., §1.4.4). One characteristic of the systems approach is the use of data-intensive functional genomics techniques: e.g., transcriptomics, metabolomics and proteomics.

Systems medicine applies systems biology to medicine. While it retains the data-driven approach to discovery that is a feature of systems biology, there are some important differences between systems medicine and systems biology.

First, systems medicine inherits from medicine a practical goal—diagnosis, prognosis and treatment—in addition to the theoretical goal of discovering pathophysiological mechanisms (Kyriakopoulou and Mulligan, 2010, p. 3). This practical goal means that, in systems medicine, causal discovery is as important as—if not more important than—mechanism discovery. (This is because, as we shall see in §3, causal models are more directly applicable to these practical ends than are mechanistic models.)

Second, the data to which systems medicine appeals is perhaps more diverse still than that considered by systems biology, because it includes, in addition to sub-cellular molecular data, higher-level clinical variables (e.g., size of tumour, sex of patient) and environmental features (describing, e.g., the origin of disease). Moreover, some researchers involved in systems medicine hope to make use of characteristics collected by personal health and fitness apps—such as number of steps walked in a day, weight and blood pressure—as well as entire medical histories collected by hospital and primary care IT systems. It is therefore clear that 'big data' plays an important role in systems medicine. Moreover, the causal discovery process also depends heavily on information about mechanisms, including social and environmental mechanisms in addition to the underlying physiological mechanisms and their malfunctioning variants. Evidence of mechanisms aids causal discovery in a variety of ways. For example, it helps to determine the direction of causation and to identify causal intermediaries. Moreover, because systems medicine appeals

to data at different levels of scale—ranging from the level of the genome to the level of populations—many of the variables in these datasets are related constitutively rather than causally (Craver, 2007). Evidence of mechanisms can help to determine which associations in the data are attributable to causal relationships and which are attributable to constitutive relationships.

Systems medicine seems to have emerged as a distinct field around 2009. Diseases tackled by large systems medicine research projects include AIDS, atherosclerosis, cervical cancer, chronic inflammatory bowel disease, colorectal cancer, fasciitis, liver cancer, lung disease, malaria, motor neurone degeneration, multiple sclerosis and tuberculosis.

This paper focusses on modelling in systems medicine. For discussions of modelling and the problem of model integration in systems biology, see O'Malley and Soyer (2012); Brigandt (2013); Green (2013) and MacLeod and Nersessian (2013).

The promise of systems medicine. Systems medicine is considered to be an exciting new paradigm for medicine, largely on account of its data-driven methodology. The use of massive amounts of data promises more robust conclusions, with fewer conclusions attributable to artefacts of the data and a larger proportion attributable to genuine connections in the sampled population. The use of big data also offers the hope of increased personalisation, with so many data points that one will be able to discover causal relationships that obtain in small subpopulations, which might otherwise be washed out in the population as a whole. This increased personalisation, in turn, offers the prospect of better-targeted treatments: treatments targeted at small subpopulations or even particular individuals, rather than at the population as a whole. Furthermore, the datasets that drive systems medicine often measure very large numbers of variables. This ability to consider so many factors at once gives systems medicine the potential to discover more complex pathophysiological mechanisms than would be discoverable by more focussed studies which concentrate only on a putative cause and effect and a few potential confounding variables.

Systems medicine clearly offers a range of opportunities. This has led some of its proponents to predict that the systems approach will quickly induce a revolution in medicine. (These bold predictions are reminiscent of those made in the early years of artificial intelligence research.) Systems medicine has been called 'P4 Medicine' in the sense that it is predictive, preventative, personalised and participatory, and many ambitious claims centre round this combination of roles. For example:

> the entire healthcare industry (from pharmaceutical companies to healthcare providers, insurance companies and medical diagnostic laboratories, etc.) will also have to transform in the years to come, possibly favoring the creation of global strategic alliances between academics, industry and administrations in order to facilitate and catalyze the arrival and development of P4 Medicine. (Sobradillo et al., 2011, p. 39.)

> We stand at the brink of a fundamental change in how medicine will be practiced. Over the next 5-20 years medicine will move from being largely reactive to being predictive, personalized, preventive and participatory (P4). Technology and new scientific strategies have always been the drivers of revolutions and this is certainly the case for P4 medicine, where a systems approach to disease, new and emerging technologies

> and powerful computational tools will open new windows for the investigation of disease. Systems approaches are driving the emergence of fascinating new technologies that will permit billions of measurements on each individual patient. … We predict that emerging technologies, together with the systems approaches to diagnosis, therapy and prevention will lead to a down turn in the escalating costs of healthcare. In time we will be able to export P4 medicine to the developing world and it will become the foundation of global medicine. The "democratization" of healthcare will come from P4 medicine. … It is evident that the business plans of every sector of the healthcare industry will need to be entirely transformed over the next 10 years (Galas and Hood, 2009, p. 1)

While some of these claims may be true, the pace of change isn't as rapid as we might be led to believe. This 10-year milestone is soon upon us and as yet there remain relatively few large-scale systems medicine research projects, let alone drastic repercussions on health care in general. To give a sense of the scale of current research, in 2015 the EU allocated roughly 36 million euros to fund around six new projects specifically in the area of systems medicine; this amounts to only about 3% of the 2014-15 budget allocated to the 'health, demographic change and wellbeing', which is itself only one of the streams of EU funding for health research.

§2
Challenges for Systems Medicine

Why is the promise of systems medicine not being realised as quickly as some have anticipated?

One reason, explored in detail by Carusi (2014), is that it can be challenging to validate models in the interdisciplinary setting of a large systems medicine project, because team members may disagree as to what counts as validation.

A second reason is simply that it is hard to handle big data. The systems approach demands a lot in terms of consistency of measurement over time and between health authorities distributed across a continent or even across the globe. Big data also make big demands in terms of computational complexity. Most algorithms for constructing a model from a dataset require computational resources that increase non-linearly with respect to the number of variables measured in the dataset. When the dataset measures several thousand variables, as can be the case with molecular-level measurements, even a quadratic-time algorithm can be computationally infeasible. Because of this issue of computational complexity, the systems approach is often forced to make a large number of simplifying assumptions, to bring the complexity down to manageable proportions. These simplifications can work against some of the advantages of the big-data approach. It is not obvious that big data together with simplifying assumptions will necessarily lead to more robust conclusions than using smaller datasets while avoiding over-simplifications.

A third—and perhaps the biggest—challenge facing systems medicine is that of data integration. How should the various datasets and other sorts of evidence combine to yield an over-arching model or set of models? Vandamme et al. (2013) introduce this challenge as follows:

> the technologies to get large amounts and different types of data will soon be affordable and readily available in the clinic. But what are we going to do with these long lists of data? Taking all this data into account, and integrating it, is not a trivial task when taking decisions in the daily practice. The sheer volume of data necessitates multidisciplinary interaction; a general practitioner cannot make diagnostic and therapeutic decisions based on hundreds of thousands of data points of -omics data by integrating it in his or her head, they require support of experts from other fields. The development of mathematical and information science tools has opened up possibilities to mine these large sets of data, to post-process them and to reduce the noise in the data… There is a need for flexible, integrative systems approaches to combine such -omics data with clinical, societal and environmental factors including sex, type of work, sleep and eat habits, etc. (Vandamme et al., 2013, pp. 892–893.)

The current approach to data integration in systems medicine often proceeds as follows (see, e.g., Lefaudeux, 2014). First, each dataset yields a 'fingerprint'. This is a model that gives an indication of the connections amongst the variables in that particular dataset. For example, a systems medicine project might produce a fingerprint model for each of the following sorts of data: metabolomic, proteomic, transcriptomic and clinical data, and patient-reported outcomes. Moreover, it is typical for most of these kinds of data to be collected both in animals (e.g., mice) as well as in humans. Ten datasets, then, would generate ten fingerprint models. Next, one or more models, involving the whole range of variables under consideration, are constructed that best fit all these fingerprint models: these are sometimes called 'handprint' models.

There are a number of difficulties with this process. First, there is no consensus as to how to generate a handprint model from a collection of fingerprint models. This tends to be done on a case-by-case basis, for example by putting the fingerprint models into a single representational scheme—e.g., Systems Biology Graphical Notation (Novere et al., 2009)—and using ingenuity to paper over the cracks which arise where the fingerprints are inconsistent. Given the scale of systems medicine projects, what is needed is a normative approach to data integration which can be applied in a systematic way, rather than ad hoc methods that appeal to intuitions which can sometimes be flawed.

The second difficulty is that a systems medicine project might aim to generate several different handprint models, using different modelling technologies and with different goals in mind for each model. There is a need to clarify the relationships amongst the models and to develop a normative approach to evidence integration for each substantially different kind of handprint model.

Third, there is a tendency—no doubt inherited from the protocols for evaluating evidence produced by the evidence-based medicine (EBM) movement (Clarke et al., 2013)—to think that the evidence to be integrated is entirely constituted by the datasets that have been collected: i.e., to think that, *if it's not a dataset then it's not evidence.* Thus the methodology is roughly:

$$\textit{datasets} \longrightarrow \textit{fingerprint models} \longrightarrow \textit{handprint models}$$

The upshot is that important evidence which does not take the form of datasets

tends to get sidelined or to be treated implicitly. While the data indeed constitute the bulk of the evidence required to ascertain the *correlations* that obtain between the variables of interest, evidence of the underlying *mechanisms* often comes from sources other than datasets, including searches of the physiological literature; past studies; individual case reports; biomedical imaging; autopsies; simulations; and the physiological knowledge of domain experts (Clarke et al., 2014a). The standard approach has hitherto involved a mixture of ignoring some such evidence and using intuition and experience to ensure that other such evidence constrains the handprint models in an appropriate way. While this sort of approach may work quite well in small medical studies, it is problematic in the systems medicine paradigm because there is simply too much of this sort of mechanistic evidence to simply eyeball it all and treat it intuitively. *All* the relevant evidence needs to be made explicit and integrated in a systematic way.

Of course, this is a big ask. These challenges are not going to be easy to meet, and this difficulty might explain why systems medicine is likely to develop at a slower pace than some have anticipated. Nevertheless, I shall suggest that we can make some useful inroads into these challenges, in the hope that systems medicine might eventually be put on a stronger footing.

§3
Kinds of Handprint Model

A first step towards clarifying the role that models play in systems medicine is to distinguish different kinds of model. Broadly speaking, four kinds of model are routinely employed in systems medicine: (I) association models, which are normally quantitative; (II) causal models, which are also normally quantitative; (III) qualitative mechanistic models; and (IV) quantitative mechanistic models.

I. Association Models. Association models are used to capture the extent to which variables measured in the datasets are predictive of one another, or of a particular target variable, such as *severity of disease.* Such a model can provide answers to questions such as: what are the main predictive factors of disease severity? If the patient exhibits factors X, how likely is a severe outcome?

In order to be used for accurate prediction, an association model needs to include the variables that are most correlated (with each other or, respectively, with the target variable). In essence this is the easiest of the four kinds of model to construct, because it suffices to capture the probability distribution over the variables included in the model. In practice, however, it is almost never the case that all variables in the handprint model are measured in the same dataset. Rather, each dataset measures a subset of the variables of interest, and there will normally be relatively few variables measured by more than one dataset. Therefore, while each dataset can be used to provide an estimate of the marginal distribution of those variables measured by that dataset, these marginal distributions constrain—rather than fully determine—the joint probability distribution over the whole set of variables in the handprint model. The key task is thus to determine and represent an appropriate joint probability distribution, from all those that satisfy the constraints imposed by the marginal distributions that are determined by the datasets and represented by the fingerprint models.

The qualitative relationships in an association model might be depicted as in Fig. 1. In this kind of undirected graph, sometimes called a *Markov network*, the

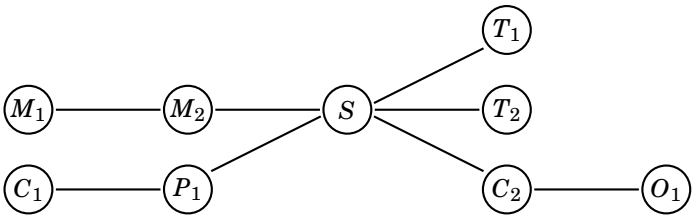


Figure 1: Qualitative representation of association relationships involving metabolic variables M_i, transcriptomic T_i, proteomic P_i, clinical C_i, patient-reported outcome O_i, severity of disease S.

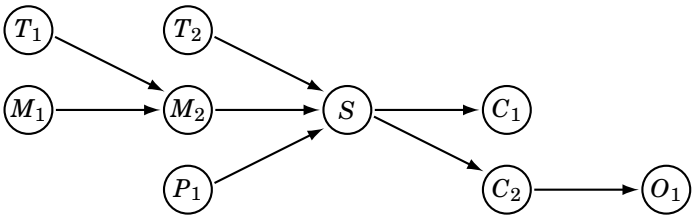


Figure 2: Qualitative representation of causal relationships involving metabolic variables M_i, transcriptomic T_i, proteomic P_i, clinical C_i, patient-reported outcome O_i, severity of disease S.

links represent correlations, and if, for sets X, Y, Z of variables, Z separates X from Y, then Z renders X and Y probabilistically independent, written $X \perp\!\!\!\perp Y \mid Z$. In Fig. 1, for instance, S separates M_1 and M_2 from T_2, so the graph implies that $\{M_1, M_2\}$ is probabilistically independent of T_2 conditional on S.

A Markov network is just one kind of probabilistic model. While association models are typically probabilistic, non-probabilistic association models, such as neural networks, also have advocates.

II. Causal Models. Causal models are similar to association models in that they model relationships between variables, including statistical associations, and can be used for prediction. In contrast to association models, however, causal models also distinguish causes from effects, usually representing causal connections graphically by means of directed acyclic graphs (DAGs) such as that depicted in Fig. 2. By explicitly representing causal relationships, causal models can be used to predict the effects of interventions: intervening to change the value of a variable will only induce further changes to those other variables in the network that are its effects, so the cause-effect relationship needs to be explicitly modelled in order to reason about interventions. Causal models thus go further than association models, in that they can be used to decide how best to control (i.e., intervene upon) variables, as well as to predict the values of certain variables when the values of other variables are observed. One increasingly common type of causal model is the causal Bayesian net (CBN), which consists of a directed acyclic graph that represents a network of causal relationships, together with the probability distribution of each variable conditional on its direct causes (Pearl, 1988, 2000).

Causal models are also used for explanation: the value that one variable is observed to take may be explained in terms of the fact that its causes take certain values which make the observed value of the effect in question more probable. This sort of explanation may be thought of a schematic abstraction of a much more nuanced explanation that describes how the underlying physiological and / or social mechanisms are responsible for the phenomenon in question.

III. Qualitative Mechanistic Models. Mechanisms may be understood in a broad sense to include physical processes (Salmon, 1984; Dowe, 2000) as well as complex-systems mechanisms, i.e., entities and activities organised in such a way as to be responsible for some phenomenon to be explained (Machamer et al., 2000; Illari and Williamson, 2012). In order to properly explain some observed phenomenon, we normally seek to describe the mechanisms that give rise to the phenomenon. For example, the progress of a disease might be explained in terms of the environmental and social processes that trigger the disease, as well as failures of the physiological mechanisms that usually protect the body from the disease, the physiological mechanisms that allow the disease to progress, and the processes of degeneration that accompany the disease. The mechanisms involved will often be hierarchically structured, involving components at the levels of society, the body, the organ, the cell, and the gene, for instance, with lower levels explaining or constituting some of the features at higher levels. Also, mechanistic explanations will typically appeal heavily to the organisation of the entities and activities involved, particularly their spatio-temporal organisation. These hierarchical and organisational features—which causal models tend to abstract away from—are explicitly represented in mechanistic models (Williamson, 2013a).

Qualitative mechanistic models often take the form of diagrams which can encapsulate these kinds of feature—see Fig. 3 for example, in which hierarchical structure and spatio-temporal organisation is clearly important.

IV. Quantitative Mechanistic Models. While a diagram is often an excellent description of the salient ingredients of a mechanistic explanation, a purely qualitative model of this form cannot fully explain why variables of interest take certain specific values—e.g., why did the disease progress for 10 years as opposed to 5 or less? In order to answer such questions, we need to introduce quantities into the model. Causal models are typically quantitative and can be used to answer questions such as this. However, as noted above, causal models explain by identifying only the key causal variables—milestones on the pathways to the effect to be explained. Sometimes such explanations are too superficial and a fuller description of the underlying mechanisms needs to be given. In such a case, a qualitative mechanistic model can be augmented with functional relationships which determine some of the quantities of interest in the mechanism in terms of others. This yields a quantitative mechanistic model. Such a model might, for example, take the form of a picture of the qualitative structure of the mechanism together with a system of differential equations linking key quantities. A second example of a quantitative mechanistic model is a recursive Bayesian net (RBN), which models a mechanism using a hierarchical array of causal Bayesian nets (Casini et al., 2011; Clarke et al., 2014b).

Another kind of mechanistic model is an agent-based model, which explains behaviour in terms of interactions between similar components, using simulations. Such models are often qualitative but can be quantitative.

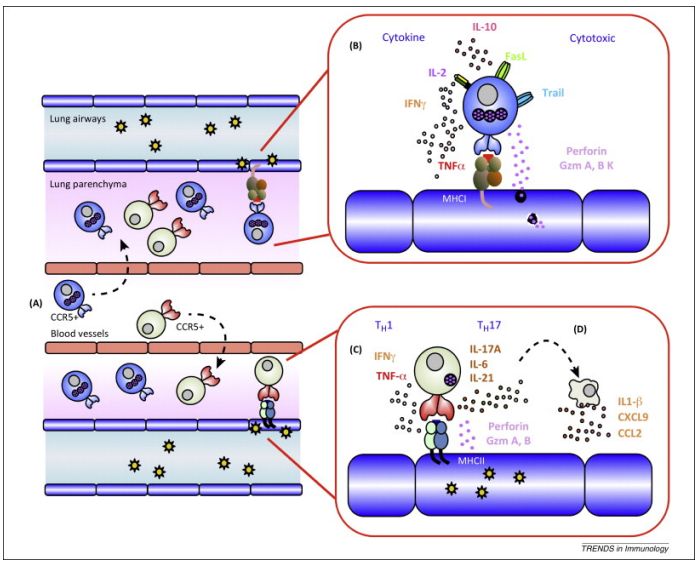


Figure 3: T cell effector mechanisms in a lung infected by influenza A virus (Gruta and Turner, 2014).

⋆

The four kinds of model are summarised in Table 1. Of these four kinds of model, the quantitative mechanistic model contains the most information and thus it is the hardest kind of model to obtain reliably. For this reason, a large-scale systems medicine project may aim to build several or all of these four kinds of model, rather than simply rely upon a quantitative mechanistic model (Wilde and Williamson, 2016). For instance, for prediction it can make much more sense to use an association model rather than a quantitative mechanistic model, because the latter kind of model will typically be far more speculative than the former kind of model, especially when evidence is limited—e.g., when evidence consists purely of datasets that record the values of the associated variables. Similarly, given limited evidence, a causal model will normally be more reliable than a quantitative mechanistic model for predicting the effects of interventions. On the other hand, a quantitative mechanistic model is the only option when certain observed quantities need to be explained in depth.

We see, then, that by classifying them according to kind of model we can begin to make sense of the array of models that are generated by systems medicine projects. There are four natural kinds of model, each of which has characteristic uses and can be used to answer distinctive questions. In §5 we shall return to the challenge of how to structure the development of models in systems medicine. In the meantime, we turn to the other key challenge facing systems medicine, that of evidence integration.

Table 1: Kinds of handprint model used in systems medicine

	Kind of model	*Kinds of question it can answer*
I.	Association	Which factors are the main predictors of brain damage severity? If the patient exhibits factors X, how likely is a severe outcome?
II.	Causal	What are the main causes of brain damage? If we intervene with drug X, how will that change the probability of a severe outcome?
III.	Qualitative Mechanistic	What explains the fact that inflammation is a cause of brain damage? What explains the fact that protein X is predictive of a severe outcome?
IV.	Quantitative Mechanistic	What explains the fact that gestational term is a better predictive factor of brain damage than MRI feature X is? What explains the fact that hypoxia doubles the chance of biomarker X?

§4
Objective Bayesian Nets as Association Models

Let us consider perhaps the main challenge facing systems medicine, namely that of evidence integration: how can one construct handprint models which take *all* the evidence into account? The standard paradigm in machine learning is to produce algorithms for constructing an association model or a causal model from a single, high-quality dataset. However, in systems medicine one is typically faced with many datasets, each pair of which may measure relatively few variables in common. In addition, there is evidence of the underlying mechanisms to take into account. How can handprint models take all this evidence into account? This section will develop a principled way of generating an association model from a range of datasets. In the next section we shall turn to the task of constructing a causal model which also takes mechanistic evidence into account.

Objective Bayesian Epistemology. The procedure for generating a handprint association model that we shall advocate in this section is motivated by a philosophical theory of strength of belief, namely objective Bayesian epistemology (OBE). Here we outline the version of OBE developed in Williamson (2010).

According to OBE, the strengths of our beliefs should satisfy three norms.

The Probability norm says that degrees of belief should be probabilities—i.e., numbers in the unit interval $[0, 1]$ such that the strength to which one believes a disjunction of disjoint propositions equals the sum of the strengths to which one believes the individual propositions. Thus, the strengths of one's beliefs should be captured by some function P in the set $\mathbb{P}$ of all probability functions.

The Calibration norm says that degrees of belief should be calibrated to evidence. In particular, degrees of belief should be calibrated to physical chances,

insofar as one has evidence of them: if it is reasonable to infer from empirical data that the chance function P^* is in some convex subset $\mathbb{P}^*$ of probability functions then one's belief function P also ought to lie in that set. Not all evidence is evidence about chances, and, more generally, the Calibration norm says that evidence will constrain P to lie in some subset $\mathbb{E}$ of probability functions that are calibrated to evidence.

The third norm—the Equivocation norm—says that, insofar as the choice of belief function P is not fully determined by the previous two norms, one's belief function should equivocate sufficiently between the most fine-grained possibilities that one can express: P should be a function in $\mathbb{E}$ that is sufficiently close to the *equivocator* function $P_=$ which gives each basic possibility the same probability. Here we shall assume that there is a finite partition Ω of basic possibilities (most fine-grained possibilities that one can express) and we shall adopt the usual measure of distance between probability functions, Kullback-Leibler divergence (KL-divergence):

$$d(P,Q) = \sum_{\omega\in\Omega} P(\omega)\log\frac{P(\omega)}{Q(\omega)},$$

for probability functions P and Q defined on Ω. If there is a function in $\mathbb{E}$ that is closest to the equivocator function $P_=$ then OBE will normally suggest that one should adopt that function. Equivalently, one's belief function P should be the probability function in $\mathbb{E}$ that has maximum entropy:

$$H(P) = -\sum_{\omega\in\Omega} P(\omega)\log P(\omega).$$

This *maximum entropy principle* was originally put forward by Jaynes (1957).

To visualise the norms of OBE, suppose there are three basic possibilities $\Omega = \{\omega_1,\omega_2,\omega_3\}$ and consider Fig. 4. The simplex $\mathbb{P}$ of all probability functions consists of the triangle linking the basic possibilities and its interior. A vertex of this triangle is the probability function that gives probability 1 to the corresponding basic possibility and probability 0 to the other two basic possibilities; edges contain the probability functions that give probability 0 to the basic possibility at the opposite vertex; interior points give non-zero probability to all three possibilities. The Probability norm says that P should lie in this triangle. The Calibration norm says that evidence eliminates all but some subset $\mathbb{E}$ of probability functions, and P should lie in this set. The Equivocation norm says that P should otherwise be closest to the equivocator function $P_=$ which gives probability $\frac{1}{3}$ to each basic possibility.

The three norms of objective Bayesianism can be motivated in terms of avoiding avoidable losses which arise when one acts or bets according to one's degrees of belief (Williamson, 2010, Chapter 3; Williamson, 2017, Chapter 9). In order to avoid the possibility of loss whichever basic possibility turns out to be true, the Probability norm must hold. In order to minimise worst-case expected loss, the Calibration and Equivocation norms must hold.

OBE for Data Integration. Consider evidence consisting of datasets $D_1,\ldots,D_k$ where each dataset D_i measures some subset V_i of the set V of variables of interest. If $V = \{X_1,\ldots,X_n\}$ then each basic possibility takes the form $X_1 = x_1,\ldots,X_n = x_n$, which we may abbreviate by $x_1\cdots x_n$.

Each dataset D_i determines a probability distribution Q_i on $V_i \subseteq V$, which tallies the frequency with which each combination of values of V_i occurs in the dataset.

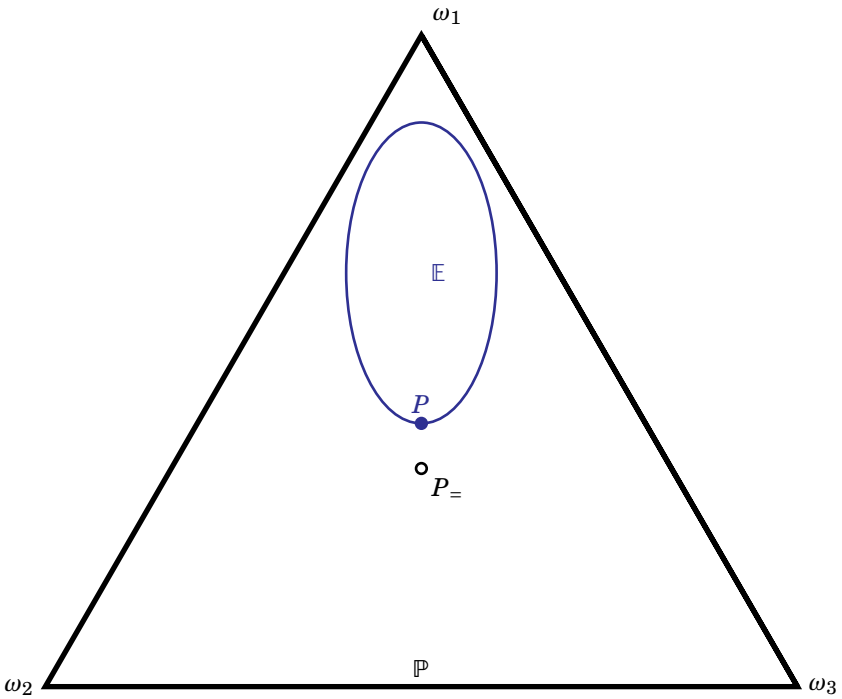


Figure 4: Visualisation of the norms of objective Bayesian epistemology.

From this data distribution, we may infer something about the data-generating chance distribution P^*. If the dataset is large enough and of sufficient quality, we may be willing to infer that $P^*_{\downarrow V_i} = Q_i$, i.e., that the chance distribution P^*, defined over the whole domain V, matches the data distribution on the subdomain V_i. In this case Q_i acts as a *point estimate* of $P^*_{\downarrow V_i}$. Otherwise, if the dataset is not sufficiently large (but still of sufficient quality), we may only be willing to infer that $P^*_{\downarrow V_i}$ lies in some convex *confidence region* around Q_i. A 95% confidence region, for instance, would be such that, if the process for generating D_i were to be repeated, in 95% of the generated datasets the induced confidence region would include $P^*_{\downarrow V_i}$. Either way, then, we have that $P^*_{\downarrow V_i}$ is constrained to lie in a closed convex set of probability functions on V_i. Note that any closed convex set of restricted probability functions on $V_i \subseteq V$ can be represented as a closed convex set of unrestricted probability functions on V, the domain as a whole. Therefore, for each dataset D_i there is some closed convex subset $\mathbb{E}_i$ of $\mathbb{P}$, defined on V as a whole, within which we infer that the chance function lies. Note that this inference is the sort of inference that is routinely drawn in classical frequentist statistics (Williamson, 2013b). It is only when calibrating a belief function P to the chances, by adopting the constraint $P \in \mathbb{E}_i$, that we move to the realm of Bayesian epistemology.

Consider a simple example. Suppose there are two binary variables, $V = \{A, B\}$. In this case the basic possibilities are $\Omega = \{ab, a\bar{b}, \bar{a}b, \bar{a}\bar{b}\}$. In order to represent this scenario in two dimensions, we shall assume that $\bar{a}\bar{b}$ is impossible, so we may restrict our attention to three basic possibilities, $\omega_1 = ab, \omega_2 = a\bar{b}, \omega_3 = \bar{a}b$. The

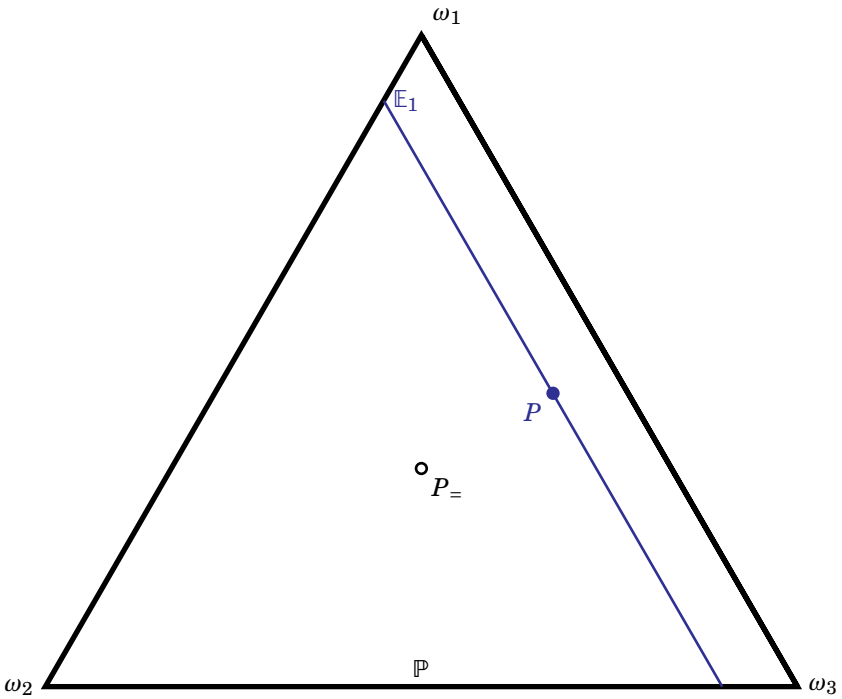


Figure 5: Single dataset, point estimate.

set $\mathbb{P}$ of probability functions is represented by the simplex $\mathbb{P}$ in Fig. 5. Suppose there is a single dataset D_1, measuring a single variable $V_1 = \{B\}$, which yields an observed frequency of 0.9 for b, $Q_1(b) = Q_1(\omega_1) + Q_1(\omega_3) = 0.9$. If we use this as a point estimate of the chance function restricted to V_1, $P^*_{\downarrow V_1}$, then the chance function must lie on the line segment $\mathbb{E}_1$ depicted in Fig. 5. Applying OBE, the Calibration norm constrains a belief function P to also lie on this line segment. The Equivocation norm fixes P to be the function on $\mathbb{E}$ closest to the equivocator function, as shown in Fig. 5. If, instead of a point estimate, we infer a confidence region $\mathbb{E}_1$, then P is yet more equivocal, as depicted in Fig. 6.

In the case of two datasets, Fig. 7 represents the objective Bayesian approach to data integration. Here we infer from dataset D_1 that P^* lies in the closed convex set $\mathbb{E}_1$ of probability functions. P_1 is the probability function that would be advocated by OBE from that dataset alone; this is the function to be represented by a fingerprint association model. Similarly for dataset D_2. In this case $V_2 = \{A\}$, and we have inferred a confidence region around the data distribution $Q_2(a) = Q_2(\omega_1) + Q_2(\omega_2) = 0.8$. Since we infer that P^* lies in $\mathbb{E}_1$ and we infer that P^* lies in $\mathbb{E}_2$, we infer that it lies in both $\mathbb{E}_1$ and $\mathbb{E}_2$, i.e., in their intersection. Thus the set of calibrated probability functions is $\mathbb{E} = \mathbb{E}_1 \cap \mathbb{E}_2$. From this we choose the most equivocal function P. This is the function we need to represent using a handprint association model.

All this assumes that the region $\mathbb{E}_1 \cap \mathbb{E}_2$ is non-empty; only then can we say that $\mathbb{E} = \mathbb{E}_1 \cap \mathbb{E}_2$. This intersection is bound to be non-empty in our original example involving two datasets measuring variable A and variable B respectively. But in

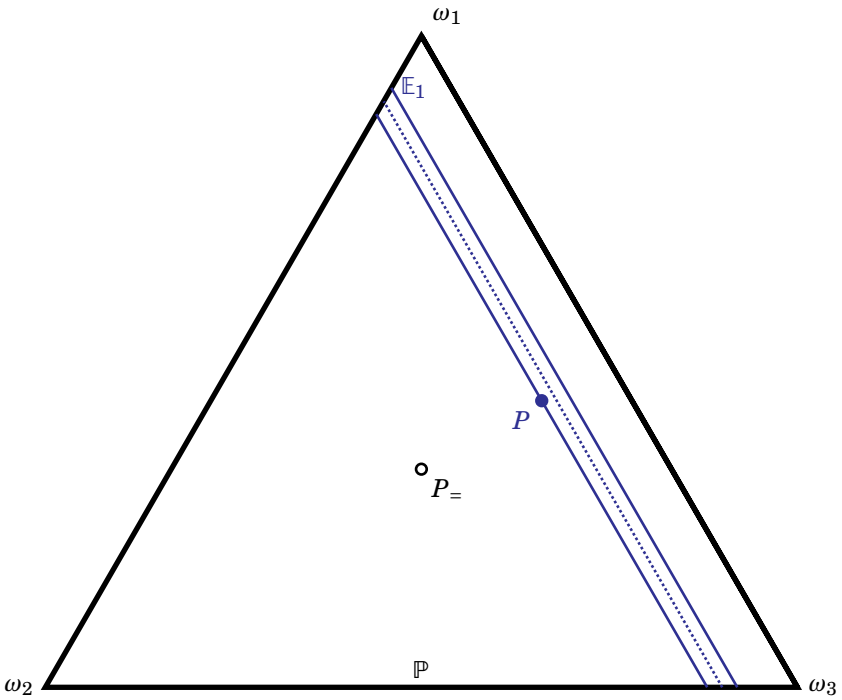


Figure 6: Single dataset, confidence region.

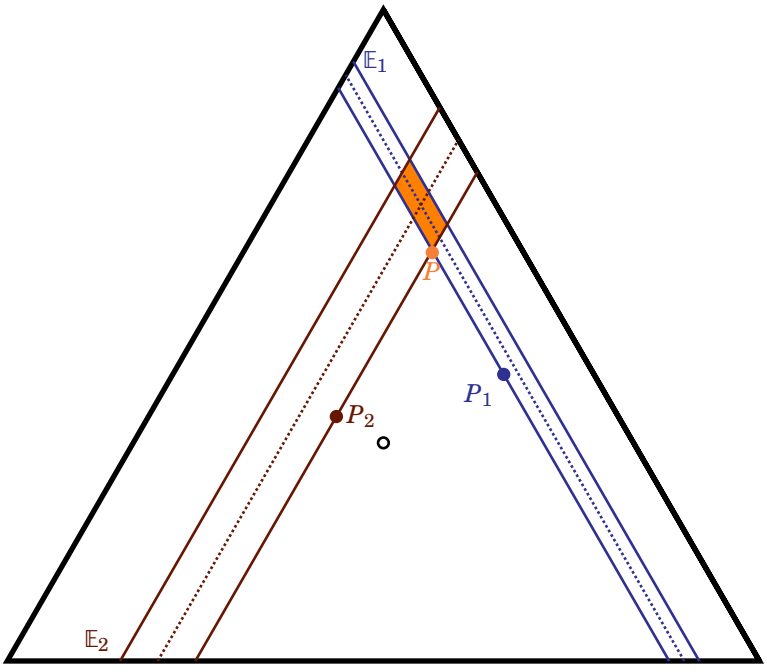


Figure 7: The objective Bayesian approach to data integration.

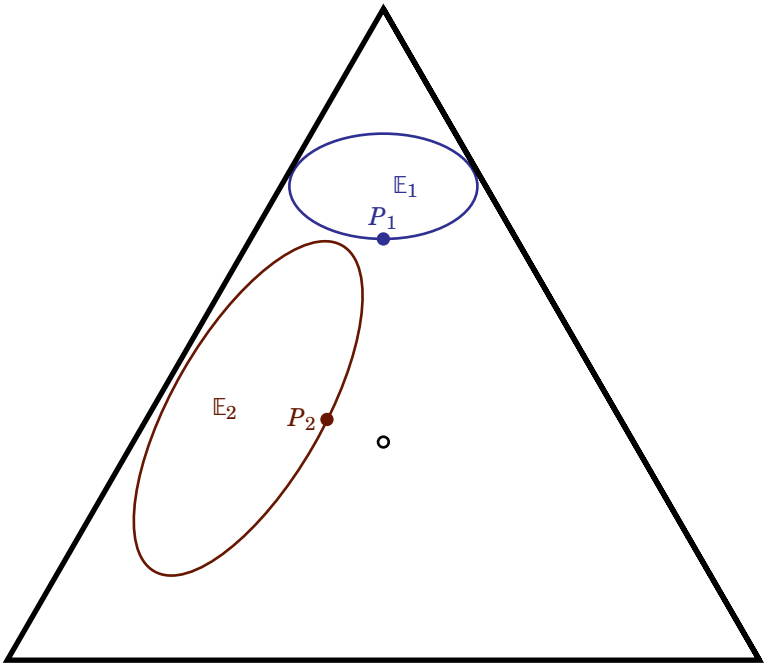


Figure 8: Inconsistent inferences: the $\mathbb{E}_i$ do not meet.

more general situations, an empty intersection, $\mathbb{E}_1 \cap \mathbb{E}_2 = \emptyset$, is a genuine possibility. If $\mathbb{E}_1$ and $\mathbb{E}_2$ do not overlap, as represented schematically in Fig. 8, then we have drawn inconsistent inferences which together imply that $P^* \in \emptyset$. The need to draw sensible conclusions about P^* motivates abandoning these original inferences and starting again, using wider confidence regions around each dataset distribution. These will translate into larger subsets $\mathbb{E}_1$ and $\mathbb{E}_2$ of $\mathbb{P}$. This process of revision needs to continue until $\mathbb{E}_1 \cap \mathbb{E}_2 \neq \emptyset$. Only then will the two inferences—namely the inference to the conclusion that P^* is in $\mathbb{E}_1$ and the inference to the conclusion that P^* is in $\mathbb{E}_2$—be consistent, and only then can we infer that P^* is in $\mathbb{E}_1 \cap \mathbb{E}_2$. This consistency maintenance process is pictured in Fig. 9. All this generalises to the situation of k datasets with which we started: from each dataset D_i one infers that the chance function P^* lies in some region $\mathbb{E}_i \subseteq \mathbb{P}$; these inferences need to be jointly consistent, and can be made so by widening the confidence regions.

Note that widening the confidence regions corresponds to *increasing* the confidence level. While the confidence level may initially be 95%, we may need to widen the regions until the confidence level is 99%, say. In Bayesian terms, we become increasingly confident that the chance function P^* is in $\mathbb{E}$. Indeed, if we are $100-\delta\%$ confident that P^* is in $\mathbb{E}_i$, for each $i = 1, \ldots, k$, then we will be at least $100-k\delta\%$ confident that P^* is in $\bigcap_{i=1}^{k} \mathbb{E}_i$; this follows from Adams' Uncertainty Theorem (Adams, 1998, Theorem 13). Therefore, a 1% increase in confidence for each region leads to at least a k% increase in confidence for the process as a whole, where k is the number of datasets. Widening the confidence regions involves a trade-off, however, in that increasing confidence often leads to a less committal belief function P. While Fig. 9 shows that increasing confidence can lead to a fairly committal belief

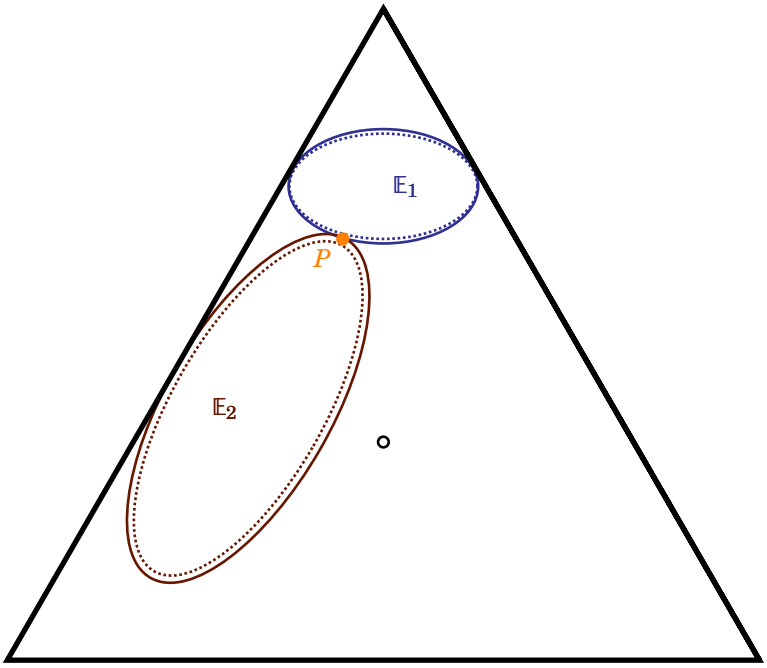


Figure 9: Widening the confidence regions to ensure that the $\mathbb{E}_i$ meet.

function, Fig. 10 depicts a case in which three dataset regions are jointly inconsistent and in which widening the regions to make them consistent would lead to the equivocator function $P_=$ as the belief function P.

This point serves to highlight the fact that the belief function P depends crucially on the chosen confidence level. Which confidence level should one choose? Plausibly, the confidence level should depend to some extent on the number k of datasets under consideration. This is to avoid situations analogous to the preface paradox: if the confidence level is 90% and there are 10 datasets then while one might infer, for each dataset, that the chance function lies inside $\mathbb{E}_i$, one would expect the chance function to lie outside one of these regions because a 90% confidence level means that on average only 9 out of 10 confidence regions would contain the chance function. Thus, it is only credible to infer that the chance function lies in the intersection of the $\mathbb{E}_i$ when the $(100-\delta)\%$ confidence level is greater than $(100-100/k)\% = 100(k-1)/k\%$. Equivalently, $\delta < 100/k$. Adams' Uncertainty Theorem gives us an even tighter bound. If $100-k\delta\% < 50\%$ then it may be more likely than not that the chance function lies outside the intersection of the $\mathbb{E}_i$. In order to avoid this possibility we need to choose $\delta \leq 50/k$. More generally, if there is a threshold τ such that we need to be at least $\tau\%$ confident in the inference that $P^* \in \bigcap_{i=1}^{k} \mathbb{E}_i$ then we need to choose $\delta \leq (100-\tau)/k$.

There is a second consideration that arises when deciding upon a confidence level. This is that one needs to balance two desiderata: confidence of inference (from the point of view of avoiding falsity, the more confident one can be about one's inferences to chances, the better) and strength of inference (from the point of view of seeking truth, the more information one can extract from the evidence,

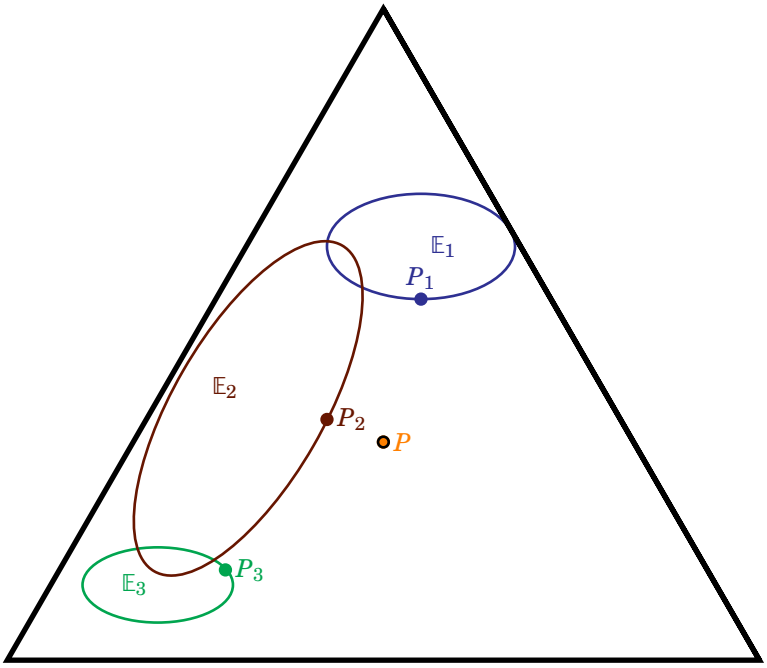


Figure 10: Widening confidence regions can lead to a less committal belief function.

the better). These two desiderata are in tension: higher confidence corresponds to wider regions which correspond to weaker inferences. The right balance will of course depend on the outcomes, such as the cost of obtaining more evidence: while weaker inferences correspond to higher confidence levels, we have seen that they can lead to a more equivocal belief function, in which case evidence is more costly insofar as it fails to sway one's beliefs. There are two extreme positions regarding the right balance between confidence and strength. One is uninteresting: if confidence absolutely trumps strength then inferences will be drawn to the widest possible confidence regions, $\mathbb{E}_i = \mathbb{P}$, representing a 100% confidence level; this will always lead to fully equivocal degrees of belief, $P = P_=$. The second extreme position is more interesting: if strength trumps confidence then use point estimates where possible—i.e., where it is consistent to do so—and otherwise increase the confidence level the minimum amount required to ensure consistency. This second position is quite plausible, for three reasons. First, although point estimates are almost always wrong, they are likely to be approximately correct—given our assumption that the data is of sufficient quality and quantity to draw these inferences in the first place—so they can still yield sensible conclusions. Second, the impact of incorrect point estimates is mitigated by the fact that inferences under OBE are highly defeasible, in the sense that as new evidence is gathered, the three norms are applied over again, so the influence of misleading evidence can soon be washed out by new inferences about chances. (This is not so under the subjective Bayesian approach, which conditionalises on total evidence and thus ensures its continuing influence.) Third, although one might worry that this extreme view leads to beliefs that are too committal, the Equivocation norm already ensures that,

given some inferred $\mathbb{E}$, degrees of belief are as equivocal as possible. Hence, opting for stronger inferences only leads to committal beliefs in a qualified sense: one is inferring as much as possible from the data, but, given that fact, degrees of belief are as equivocal as possible.

We thus have some motivation for starting with point estimates and increasing confidence levels just enough to ensure consistency. As an alternative approach, one might suggest the following recipe: instead of starting with point estimates, start with confidence regions, where the confidence level incorporates the bound arising from Adams' Uncertainty Theorem, i.e., start off with $\delta = (100 - \tau)/k$ and decrease δ just enough to ensure consistency. One should note, however, that this modification can significantly tip the balance away from strength of inference, towards confidence of inference, and can lead to very equivocal belief distributions. Consider the case where there is a large number k of datasets being integrated, none of which contains a large number of sampled individuals. In this case δ will be very small because k is large, the confidence level will be very high, and each region $\mathbb{E}_i$ will be large because of this high confidence level and the small number of observations in each dataset. Thus the intersection of these regions can be large. Moreover, if each region is large enough to include the equivocator function $P_=$ then that is the function that will be chosen as the resulting belief function. (This is the same belief function that would be chosen if there were no data at all.) It seems that much of the useful information contained in the data is being overlooked. Starting with point estimates, as suggested above, avoids this problem.

Thus far we supposed that each dataset is of sufficiently high quality to permit an inference from the dataset to a point estimate or a confidence region estimate. This is reasonable in the context of a systems medicine project, which is responsible for collecting its own data, and for its own quality control. (This assumption may be less reasonable in the context of a systematic review or meta-analysis of studies produced by disparate research teams.) Nevertheless, even within a project, different datasets can be judged to be of differing quality. What can be concluded when the datasets are of variable quality? One possibility is to appeal to the following Bayesian approach which factors in judgements of quality. Let Q_i be the proposition that dataset D_i is of high enough quality for one to be willing to infer a confidence region estimate from the data at the $100-\delta\%$ level. Then,

$$P(P^* \in \mathbb{E}_i) = P(P^* \in \mathbb{E}_i \mid Q_i)P(Q_i) + P(P^* \in \mathbb{E}_i \mid \neg Q_i)P(\neg Q_i).$$

From a Bayesian point of view, $P(P^* \in \mathbb{E}_i \mid Q_i)$ is just the confidence level; $P(Q_i)$ is the judged quality level of the dataset; and $P(P^* \in \mathbb{E}_i \mid \neg Q_i)$ might be estimated to be the proportion of the simplex contained in region $\mathbb{E}_i$. Let $x = P(Q_i)$. Then,

$$P(P^* \in \mathbb{E}_i) = \frac{100-\delta}{100}x + \frac{|\mathbb{E}_i|}{|\mathbb{P}|}(1-x).$$

This relativises $P(P^* \in \mathbb{E}_i)$ to the quality of the data. One can vary the confidence level $P(P^* \in \mathbb{E}_i \mid Q_i)$ from dataset to dataset to ensure a uniform value $P(P^* \in \mathbb{E}_i)$ across the datasets.

$\star$

We see, then, that objective Bayesian epistemology provides a systematic way of integrating datasets, each of which measures only a subset of variables of interest.

The next step is to *model* the relevant probability distributions—i.e., to represent these probability distributions efficiently, in such a way that probabilities of interest may easily be inferred from the model. Each dataset determines a probability distribution P_i; a model of such a distribution is a fingerprint model. A model of the distribution P which integrates all the data is a handprint model. (These models are all association models—they can be used for prediction, but they do not model causal relationships or mechanisms. Causal and mechanistic models will be considered separately in §5.) Recall from §2 that the standard approach in systems medicine is to determine a handprint model directly from the fingerprint models. Fig. 7 gives us reason to question this approach: while the integrating probability distribution P is determined by that data regions $\mathbb{E}_i$, it does not seem to be determined by the individual distributions P_i—there does not seem to be enough information encapsulated in these regions to determine P.

Next we shall propose a particular kind of fingerprint and handprint association model, namely an *objective Bayesian net* model. We shall see that at least in some cases, the fingerprint models do determine the handprint model.

Objective Bayesian nets. Bayesian nets have become perhaps the model of choice for representing and reasoning with a probability distribution on a finite number of discrete variables. This is because, while probabilistic inference is extremely computationally complex in the worst case, probabilistic inference using Bayesian nets tends to be quite efficient in typical cases. Indeed, a wide variety of typically efficient algorithms have been developed, both for constructing Bayesian nets (e.g., Neapolitan, 2004) and for inferring probabilities from them (e.g., Darwiche, 2009).

A Bayesian net consists of a directed acyclic graph (DAG) with the variables as nodes, together with the probability distribution of each variable conditional on its parents in the DAG. The main modelling assumption is the *Markov Condition*: each variable is probabilistically independent of its non-descendants in the graph, conditional on its parents. Under this assumption, a Bayesian net determines a joint probability distribution over the set of variables under consideration.

An *objective Bayesian net* (OBN) is a Bayesian net which represents a probability distribution advocated by OBE, i.e., a probability distribution which fits available evidence but which otherwise has maximum entropy. Thus OBNs can be used to represent the functions P_i and P outlined above. OBN representations of the P_i constitute fingerprint association models, while an OBN representing P is a handprint association model.

Let us consider how one might apply OBNs to data integration in systems medicine. We shall sketch the simplest case, where there are k consistent data distributions $Q_1,\ldots,Q_k$, and these are used as point estimates, i.e., one infers that $P^*_{\downarrow V_i} = Q_i$ for $i = 1,\ldots,k$. This case is treated in more detail in Landes and Williamson (2016). See Williamson (2005b) for a general introduction to OBNs.

The first task is to model each data distribution Q_i. This can be done by applying the standard machine learning techniques alluded to above, to construct a Bayesian net that represents the data distribution in question. Note that this Bayesian net only involves variables in V_i. Since we infer that the chance distribution matches the data distribution on V_i, $P^*_{\downarrow V_i} = Q_i$, and the Calibration norm says that one's belief function should match the chance distribution insofar as evidence determines the chance distribution, this belief function should match the data distribution, $P_{i\downarrow V_i} = Q_i$. Thus the Bayesian net model of Q_i can also be thought of

as a model of the probability distribution P_i advocated by OBE given evidence solely consisting of dataset D_i. Consequently, this Bayesian net model of Q_i is a fingerprint model.

The second task is to build a handprint OBN model, which represents the distribution P advocated by OBE given all the datasets $D_1,\ldots,D_k$. In order to build the DAG in the model, first construct an undirected graph on V by linking each pair of variables that occur in the same subdomain V_i, for any $i = 1,\ldots,k$. This undirected graph is a Markov network (§3) which represents probabilistic independencies that are guaranteed to be satisfied by P: if $Z \subseteq V$ separates $X \subseteq V$ from $Y \subseteq V$ in this graph then X and Y are probabilistically independent conditional on Z, $X \perp\!\!\!\perp_P Y \mid Z$, for the rational belief function P (Williamson, 2005a, Theorem 5.1). (That these independencies provably hold is attributable to the fact that P maximises entropy. Of course it may be that the chance function P^* does not satisfy all these independencies, but as the data do not provide any information as to whether this is the case or not, the best one can do is act in accordance with the norms of OBE, i.e., in accordance with the rational belief function P.) Now P will satisfy further probabilistic independencies: for each i, $P_{\downarrow V_i} = Q_i$ and the Bayesian net model of Q_i implies certain independencies which cannot be inferred from the undirected graph by means of the separation criterion. In order to capture as many of these independencies as possible, one can prune edges from the undirected graph to yield a sparser Markov network.[1] Next, one can transform the resulting sparser undirected graph into a DAG such that the Markov Condition is bound to hold—an algorithm for doing so is presented in Williamson (2005a, §5.7). This is the DAG in the OBN representation of P. It remains to specify the probability distribution of each variable conditional on its parents in the graph. Some of these distributions can be obtained very straightforwardly (Landes and Williamson, 2016). For example, if a variable and its parents all occur in the same subdomain V_i for some i, then the distribution can be obtained from the corresponding fingerprint model. In other cases, an optimisation problem must be solved in order to find the maximum entropy solution.

This procedure is illustrated in Fig. 11 in the case of two datasets, each of which measures three binary variables. In this illustration, the DAG of the handprint model looks much like the composition of the DAGs of the fingerprint models. While this is sometimes so, it is not always the case. However, in our set-up of k consistent point-estimate data distributions $Q_1,\ldots,Q_k$, it is always the case that the Bayesian net fingerprint models determine the handprint model. This provides a partial vindication of the systems medicine methodology of constructing a handprint model from fingerprint models. It is only a partial vindication because this does not apply to the more general approach involving confidence region estimates rather than point estimates. There, the handprint model will depend on the confidence regions themselves, not simply on the fingerprint model: the fingerprint models do not 'screen off' the data from the handprint model.

Having sketched a principled way of integrating data to form a handprint association model, we shall now turn to the task of constructing a handprint causal

[1] This can be done as follows. The undirected graph constructed above can be thought of as $\bigcup_{i=1}^k \mathscr{C}_i$, where $\mathscr{C}_i$ is the complete graph on V_i. Instead of this graph, one can take a sparser graph in two steps. First take $\bigcup_{i=1}^k \mathscr{M}_i$, where $\mathscr{M}_i$ is the *moral graph* on V_i formed by (i) taking the DAG in the Bayesian net model of Q_i, (ii) for each variable in V_i joining each pair of its parent variables by an edge, and (iii) finally removing the orientations on the remaining arrows. Second, ensure that each pair of variables that occur together in more than one variable set are connected by an edge.

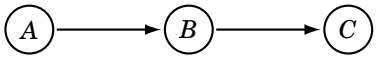
Dataset D_1:

	A	B	C
Jill	✓	✗	✓
Keith	✓	✗	✗
Linda	✗	✗	✓
…	…	…	…

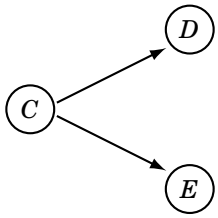
Dataset D_2:

	C	D	E
Jim	✗	✗	✓
Kirsty	✓	✓	✓
Lionel	✗	✓	✗
…	…	…	…

Bayesian net fingerprint model 1:



Bayesian net fingerprint model 2:



Objective Bayesian net handprint model:

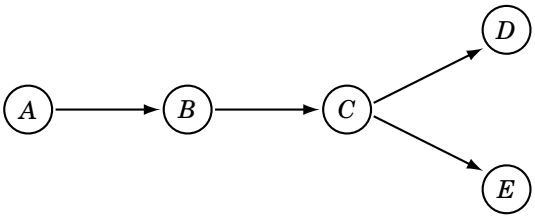


Figure 11: An OBN handprint model from Bayesian net fingerprint models.

model.

§5
Causal and Mechanistic Models

In this section we shall outline a standard methodology for constructing a causal model, argue for an alternative methodology, and develop a general framework for model construction in systems medicine.

Standard methodology. As mentioned in §3, one common kind of causal model is a *causal Bayesian net* (CBN). This is a Bayesian net whose DAG conveys information about causal connections as well as probabilistic independencies: an arrow from one variable to another indicates that the former variable is a direct cause of the latter. In the CBN case, the Markov Condition, now called the *Causal Markov Condition*, says that each variable is probabilistically independent of its non-effects, conditional on its direct causes.

The standard approach to generating a CBN is a data mining approach: input a dataset; find a Bayesian net which best fits the data distribution (or a class of nets which best fit the data distribution); interpret the arrows in the DAG (or those

arrows common to all DAGs in the class of nets) as direct causal connections or as indicative of unmeasured common causes (Spirtes et al., 1993; Gammerman, 1999; Glymour and Cooper, 1999; Pearl, 2000, Chapter 2). According to this standard approach, causal discovery can be automated by implementing appropriate data mining algorithms. This approach has been so influential that causal learning is now considered to be one of the subfields of machine learning.[2] This standard approach is thus closely related to the way in which non-causal Bayesian networks are constructed from data. The idea is roughly that the arrows in such a Bayesian net represent a pattern of statistical associations and independencies, and that the best explanation of this pattern is that there are corresponding causal connections which give rise to the pattern, so we may interpret the arrows causally.

This standard data mining methodology needs to be adapted to fit the systems medicine context. This is because, in the systems medicine context, there is not a single dataset on the whole domain V from which to learn a causal handprint model—there are lots of datasets on various subdomains V_i. Thus, while one may be able to apply the standard methodology to yield causal fingerprint models, further methods are required in order to generate a handprint model. One such method was developed by Danks (2002); Tillman et al. (2008) and Tillman and Spirtes (2011): this involves patching together several CBN fingerprint models in order to generate a CBN handprint model.

The main difficulty with both this method and the standard data mining methodology is that they assume that all correlations should be explained causally. Even in the case of the fingerprint models, a causal interpretation of the arrows in the network is controversial (e.g., McKim and Turner, 1997) and indeed often implausible. The best explanation of an association in the data may not be a corresponding causal connection. In many cases the best explanation is measurement error or bias; or that the variables are associated in virtue of being time-series measurements, or in virtue of a semantic, constitutive, logical, mathematical or physical connection rather than a causal connection (Williamson, 2005a, §4.2). As medicine knows to its cost, only very rarely does a new association in data turn out to be genuinely causal.

Indeed, at least in the health sciences, establishing association is not normally enough to establish causation. Typically, in order to establish that A is a cause of B, one needs to establish not only that A and B are appropriately associated (i.e., that they are probabilistically dependent conditional on B's other direct causes), but also that there is some underlying mechanism linking A to B which can account for this association and by which one can explain instances of B by citing instances of A (Russo and Williamson, 2007). Only then is the best explanation of the association between A and B that A is a cause of B.

This observation suggests a revision to the standard methodology for generating a CBN: one should base a causal model on evidence of mechanisms as well as on the associations found in data. Evidence of mechanisms tends to be much more multifarious than evidence of association. For instance, evidence of mechanisms can be had by manipulation (e.g., in vitro experiments), by observation (e.g., biomedical imaging, autopsy, case reports), from statistical trials (e.g., randomised controlled trials), from confirmed theory, by analogy (e.g., animal experiments), and by simulation (e.g., agent-based models). Clarke et al. (2014a) argue that all such

[2] Challenge problems are sometimes devised to test causal discovery algorithms—see, e.g., Causality Workbench, http://www.causality.inf.ethz.ch/.

evidence needs to be taken into account when establishing causal claims. As noted in §1, evidence of mechanisms can help to distinguish those correlations that are causal from those that are attributable to other considerations, such as constitutive relationships.

Alternative methodology. This sort of revised methodology can be developed as follows. The goal is to generate CBNs that best fit both the pattern of dependencies and independencies suggested by the data and the causal constraints imposed by evidence of underlying mechanisms. Instead of generating regular CBNs, however, we shall consider *labelled* CBNs, that is, CBNs whose arrows are annotated. The label attached to an arrow provides information about the kind of connection represented by the arrow. Some arrows will be labelled as causal, while others may signify constitutional, semantic or logical connections, for example, and labelled as such.

For instance, mechanistic evidence might establish that A is not a cause of B (perhaps because A only occurs after B, or because A is related to B constitutionally rather than causally), in which case it imposes the following constraint on a causal handprint model: there should be no chain of causal arrows (i.e., arrows that are labelled as causal) from A to B. If the mechanistic evidence determines that A is related to B constitutionally rather than causally, then there is a further constraint that A and B should be connected in the DAG by arrows that are labelled as constitutional.

The key task is then to identify those labelled CBNs which (i) satisfy the constraints imposed by mechanistic and other evidence, (ii) explain all the dependencies in the OBN by labelled arrows in the CBN, and (iii) posit as few connections as possible that do not correspond to dependencies in the OBN. One practical method for carrying out this task will be developed shortly.

The main differences between this approach and the standard CBN data mining methodology are as follows. First, evidence other than the data is taken into account by condition (i). Second, the standard CBN methodology explains all the dependencies in the data by *causal* relationships; here, condition (ii) requires merely that every dependency be explained in *some* way, by invoking a causal, mechanistic, semantic, or some other relationship. (In certain cases it may be that, given all available evidence, the best explanation of a dependency in the data is that it is accidental, attributable for example to a small sample size. In such cases, the corresponding arrows may be labelled 'accidental'.)

If necessary, the labels attached to the arrows of the CBN can be further refined. In particular, once the labelled DAG is constructed, one can evaluate each of the causal claims posited by the model, considering a causal relationship to be *established* just when the available evidence establishes both an appropriate association (with respect to the chance function P^* rather than the rational belief function P) and that this association is causal, i.e., that there exists some suitable mechanism which appeals to the putative cause to explain the putative effect. One can then further label the corresponding arrow in the DAG as 'established'. Those causal claims that cannot be considered to be established may be classified as *provisionally established* (or, more simply, *provisional*); *arguable*; or *speculative*, ordered according to increasing likelihood that future evidence will lead to such claims being revisited and rejected (Parkkinen et al., 2017). This leads to a more fine-grained classification of causal relationships.

Having constructed the labelled DAG of the CBN, it remains to specify the probability distribution of each variable conditional on its parents in the DAG. The joint probability distribution P, defined over the domain as a whole, is already fully determined by the OBN, and the required marginal distributions can thus be obtained directly from the OBN by applying standard Bayesian network inference algorithms.

Constructing a labelled DAG. So, how should the labelled DAG be constructed in practice? We shall describe an approach which presupposes that the available evidence fully determines the non-causal relationships. It is often the case that mechanistic evidence determines the constitutional relationships but does not fully determine the causal relationships. Similarly, evidence of semantic relationships among the variables may well fully determine which variables have overlapping meaning yet fail to determine, of those variables that represent disjoint events, which causes which. If it is indeed the case that the non-causal relationships amongst the variables are known, then any unexplained correlations in the data are best explained causally. If so, then, by default, causal relationships should be posited to explain unexplained correlations.

The first step is to construct a model which best represents those DAGs that chart the probabilistic independencies of the joint probability distribution P. The standard way to represent a *Markov equivalence class* of DAGs—i.e., a class of DAGs that characterise the same set of independencies—is to construct a *partially directed acyclic graph* (PDAG). This is a graph which may contain a mix of arrows and (undirected) edges, with no directed cycle. A PDAG represents an equivalence class of DAGs when every completion of the PDAG (an orientation of the undirected edges in the PDAG which produces a DAG) lies in the equivalence class, and, vice versa, every member of the class is a completion of the PDAG.

This PDAG is straightforward to construct, given a catalogue of the independencies of the probability distribution. Koller and Friedman (2009, §3.4), for example, provide simple algorithms for generating the PDAG which represents the equivalence class of minimal DAGs that characterise the independencies of a joint distribution. These independencies can be read off the undirected Markov network which was constructed in §4 as a stepping-stone to an OBN: recall that, if Z separates X from Y in this graph then X and Y are probabilistically independent conditional on Z.[3]

Having constructed the PDAG, the second step is an iterative approach to solving a constraint satisfaction problem. The DAGs represented by the PDAG satisfy condition (iii) given above: they posit as few connections as possible that do not correspond to dependencies in the OBN, since they are minimal models of the independencies of P. It remains to satisfy the other conditions: i.e., to find the labelled DAGs that satisfy the constraints imposed by mechanistic and other evidence. This can be done by searching through the DAGs represented by the PDAG, seeing whether some labelling exists that satisfies the constraints, and rejecting those DAGs that do not satisfy the constraints. If no such minimal DAG satisfies the constraints, then we can proceed to search incrementally through DAGs with

[3] If all the independencies of P can be represented by some DAG then the PDAG is uniquely determined. (A sufficient condition for this is that the undirected graph used in the construction of the OBN is triangulated.) If not, then a class of PDAGs will be needed to represent the class of all possible DAGs compatible with P (see, e.g., Koller and Friedman, 2009, §3.4).

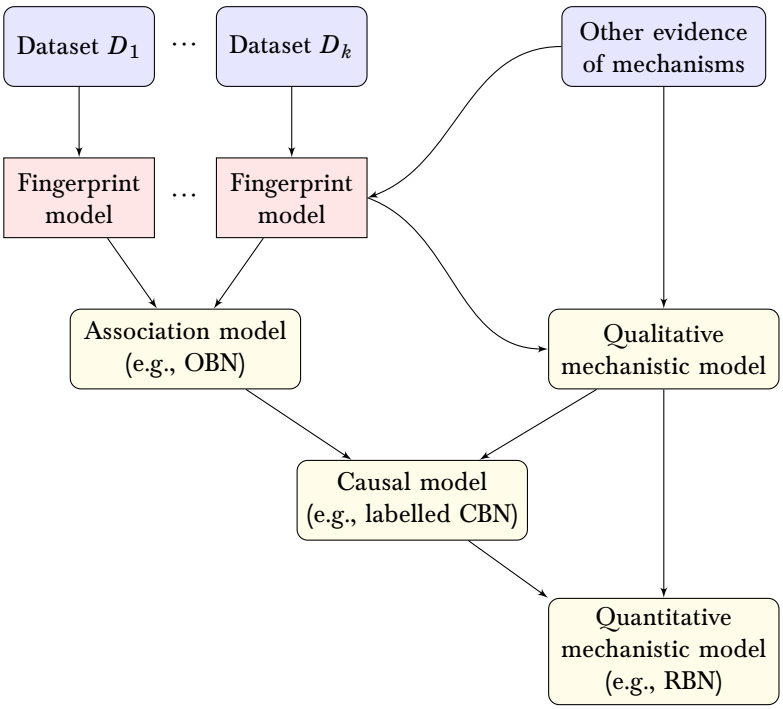


Figure 12: Relationships amongst models.

one or more extra arrows. Assuming the constraints are satisfiable at all, this culminates in a set of minimal labelled CBNs that satisfy the constraints imposed by *all* the evidence—not just the datasets.

The more fine-grained labelling might be used to select certain preferred models from within this set for the purpose of inference. The aim would be to select those causal handprint models whose causal relationships are less prone to revision—i.e., to prefer established causal relationships over those that are provisional, provisional over arguable, and arguable over speculative. The handprint model can then be used for causal inference—e.g., for deciding which of the correlates of an outcome variable, such as severity of disease, to intervene upon in order to alleviate the disease.

General framework for model construction. In terms of the array of handprint models discussed in §3, the causal model depends not only on the correlations represented by an association model, but also on the mechanisms represented by a qualitative mechanistic model. The general picture is presented in Fig. 12. This depicts the situation in which the fingerprint models screen off the data from the association model, as is the case for instance with point-estimate Bayesian net fingerprint models and an OBN handprint model. Note that evidence other than data can influence the choice of fingerprint models. For example, information about mechanisms assists with the design and interpretation of statistical trials (Clarke et al., 2014a). Similarly, associations found in the data and represented in the fingerprint models can suggest new mechanistic connections, and so influence the development of a

qualitative mechanistic model. As suggested above, a causal handprint model needs to be influenced by both the associations represented in the association handprint model and the pattern of mechanistic connections implied by a qualitative mechanistic model. This mechanistic model may well be tentative in parts or incomplete. Causal models tend to be quantitative, so the causal model can, in turn, influence the development of a quantitative mechanistic model by suggesting relationships amongst quantities that feature in the mechanism.

It is thus possible to develop a principled methodology for structuring the development of models in systems medicine—an array of models which can otherwise seem bewildering to those involved in a large systems medicine project, let alone those outside the project trying to comprehend its results. As a further simplification, the Bayesian network modelling formalism can be used to unify the models employed in such a project: standard Bayesian nets can act as useful association fingerprint models; an objective Bayesian net constitutes an association handprint model; as we have just seen, in conjunction with qualitative information about mechanisms, such a net is a natural stepping stone to a labelled causal Bayesian net handprint model; finally, recursive Bayesian nets (see §3) allow one to draw inferences across the levels of a hierarchical mechanism, and so act as a kind of quantitative mechanistic model.

While this approach structures and unifies modelling in systems medicine, it does not make modelling easy. In particular, it is harder to automate the discovery of causal relationships under this approach than it is under the standard (but problematic) data-mining approach. This is because causal claims need to track mechanistic connections as well as associations in the data, and thus the whole range of evidence of mechanisms needs to be evaluated in order to determine how it constrains the causal claims that can be made. Therefore, the pathways from evidence to a qualitative mechanistic model, depicted in Fig. 12, are not readily automated. Moreover, it is not easy to automate the move from qualitative mechanistic models that take the form of diagrams or pictures, such as Fig. 3, to constraints on a CBN. In order to automate this step, one needs other forms of qualitative mechanistic model which are more accessible from a computational point of view (though invariably less intuitive to humans). These sorts of models, standardised by being represented in Systems Biology Markup Language (SBML) and structured in terms of semantic relationships specified in various biomedical ontologies, are increasingly prevalent in systems medicine (see, e.g., Hoehndorf et al., 2011).

§6
Conclusions and Open Questions

Progress in systems medicine has been slower than anticipated. Arguably this for two main reasons: (i) evidence integration is a challenge; (ii) building and comprehending a large array of models is a challenge. In this paper I have tried to take a constructive approach to these two challenges, by setting out a possible modelling methodology for systems medicine. At a general level, this methodology seeks to elucidate the relationships between models, as well as to structure their development, as depicted in Fig. 12. At a more specific level, this methodology fits very well with the Bayesian net approach to modelling, and objective Bayesian nets offer a natural framework for data integration.

This paper has focussed on *theoretical models* in systems medicine—models that are abstractions of the phenomena they seek to model. *Biological models*—i.e.,

models that are organisms or parts of organisms (Wilde and Williamson, 2016)—are also widely applied in systems medicine. While the relationship between theoretical and biological models is somewhat complex, an interesting next step would be to integrate the role of biological models into this modelling framework.

In terms of developing the OBN formalism, while OBNs are currently well understood in the case of consistent point-estimate data distributions, this is far less so in the case of confidence-region estimates. This is one natural direction for further research. Another direction involves extending the OBN methods introduced above, which deal with datasets that measure subsets of V, i.e., *unconditional* marginal distributions of P^*, to those that measure *conditional* marginal distributions. Often, a study will examine only a subpopulation of the target population, e.g., patients who are not pregnant, and its conclusions can be thought of as pertinent to estimating a conditional chance distribution, e.g., $P^*(\cdot|\neg Pregnant)$. While the objective Bayesian principles of data integration that are outlined above extend naturally to this situation, it remains to be seen how the algorithms for OBN construction should best be adapted to the conditional distribution case.

With regard to the labelled CBN formalism, two tasks are particularly pressing. First, the approach introduced above dealt with the case in which any unexplained correlation is to be explained causally. A strategy is needed for dealing with other situations—e.g., those in which one needs to decide whether to introduce a causal claim or a constitutional claim to explain a correlation. One strategy is to appeal to Craver's characterisation of constitution relations in terms of mutual manipulability (Craver, 2007). However, this approach has its detractors (e.g., Leuridan, 2012; Baumgartner and Gebharter, 2015), and bears closer scrutiny. Second, predicting the effects of interventions can be non-trivial when some of arrows in the model are non-causal. In a CBN in which all the arrows are causal, one can predict the effect of interventions by deleting the arrows that lead into the variable which is intervened upon, and then using standard Bayesian network inference algorithms to update the probabilities of variables of interest. In the labelled CBN formalism, if some of the arrows incident upon the intervention variable are non-causal, it can be far from obvious as to which—if any—of these should be deleted. The prospects of a generalised approach to intervention is an interesting question for further research.

Acknowledgements

I am grateful to the UK Arts and Humanities Research Council and the Leverhulme Trust for funding this research and to Jürgen Landes, Veli-Pekka Parkkinen, Federica Russo, Christian Wallmann, Michael Wilde and two anonymous referees for helpful comments.

Bibliography

Adams, E. W. (1998). *A primer of probability logic*. CSLI Publications, Stanford.

Baumgartner, M. and Gebharter, A. (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science*, Online first.

Boogerd, F. C., Bruggeman, F. J., Hofmeyr, J.-H. S., and Westerhoff, H. V., editors (2007). *Systems Biology: Philosophical Foundations*. Elsevier, Amsterdam.

Brigandt, I. (2013). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4A):477–492.

Carusi, A. (2014). Validation and variability: Dual challenges on the path from systems biology to systems medicine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 48:28–37.

Casini, L., Illari, P. M., Russo, F., and Williamson, J. (2011). Models for prediction, explanation and control: recursive Bayesian networks. *Theoria*, 26(1):5–33.

Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventative Medicine*, 57(6):745–747.

Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2014a). Mechanisms and the evidence hierarchy. *Topoi*, 33(2):339–360.

Clarke, B., Leuridan, B., and Williamson, J. (2014b). Modelling mechanisms with causal cycles. *Synthese*, 191(8):1651–1681.

Craver, C. F. (2007). *Explaining the brain*. Oxford University Press.

Danks, D. (2002). Learning the causal structure of overlapping variable sets. In Lange, S., Satoh, K., and Smith, C. H., editors, *Discovery Science: Proceedings of the 5th International Conference*, pages 178–191, Berlin. Springer.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. Cambridge University Press, New York.

Dowe, P. (2000). Causality and explanation: review of Salmon. *British Journal for the Philosophy of Science*, 51:165–174.

Galas, D. J. and Hood, L. (2009). Systems biology and emerging technologies will catalyze the transition from reactive medicine to predictive, personalized, preventive and participatory (P4) medicine. *Interdisciplinary Bio Central*, 1(6):1–5.

Gammerman, A., editor (1999). *Causal models and intelligent data management*. Springer, Berlin.

Glymour, C. and Cooper, G. F., editors (1999). *Computation, causation, and discovery*. MIT Press, Cambridge MA.

Green, S. (2013). When one model is not enough: Combining epistemic tools in systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4A):170–180.

Gruta, N. L. L. and Turner, S. J. (2014). T cell mediated immunity to influenza: mechanisms of viral control. *Trends in Immunology*, 35(8):396–402.

Hoehndorf, R., Dumontier, M., Gennari, J. H., Wimalaratne, S., de Bono, B., Cook, D. L., and Gkoutos, G. V. (2011). Integrating systems biology models and biomedical ontologies. *BMC Systems Biology*, 5(124):1–16.

Illari, P. M. and Williamson, J. (2012). What is a mechanism? thinking about

mechanisms *across* the sciences. *European Journal for Philosophy of Science*, 2:119–135.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models*. MIT Press, Cambridge, MA.

Kyriakopoulou, C. and Mulligan, B. (2010). From systems biology to systems medicine. European Commission, DG Research, Directorate of Health workshop report. ftp://ftp.cordis.europa.eu/pub/fp7/health/docs/final-report-systems-medicine-workshop_en.pdf.

Landes, J., Osimani, B., and Poellinger, R. (2017). Epistemology of causal inference in pharmacology: Towards a framework for the assessment of harms. *European Journal for Philosophy of Science*, in press.

Landes, J. and Williamson, J. (2016). Objective Bayesian nets from consistent datasets. In Giffin, A. and Knuth, K. H., editors, *Proceedings of the 35th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1757 of *American Institute of Physics Conference Proceedings*, Potsdam, NY.

Lefaudeux, D. (2014). U-BIOPRED toolbox for fingerprint and handprint generation. Biomax Symposium 2014: 'Translating systems medicine into practice'. http://www.biomax.com/symposium/2014/201409_lefaudeux.pdf.

Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science*, 63(2):399–427.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67:1–25.

MacLeod, M. and Nersessian, N. J. (2013). Coupling simulation and experiment: The bimodal strategy in integrative systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4A):572–584.

McKim, V. R. and Turner, S. (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*. University of Notre Dame Press, Notre Dame.

Neapolitan, R. E. (2004). *Learning Bayesian networks*. Pearson / Prentice Hall, Upper Saddle River NJ.

Novere, N. L., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villeger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T. C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., and Kitano, H. (2009). The systems biology graphical notation. *Nature Biotechology*, 27(8):735–741.

O'Malley, M. A. and Soyer, O. S. (2012). The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1):58–68.

Parkkinen, V.-P., Wallmann, C., Wilde, M., Williamson, J., Clarke, B., Illari, P., Kelly, M., and Russo, F. (2017). Evaluating evidence of mechanisms in medicine: a handbook for practitioners. http://ebmplus.org/handbook-practitioners-beta-version/. Version 0.1.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible*

inference. Morgan Kaufmann, San Mateo CA.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge.

Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2):157–170.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press, Princeton NJ.

Sobradillo, P., Pozo, F., and Agusti, A. (2011). P4 medicine: the future around the corner. *Archivos de Bronconeumologia*, 47(1):35–40.

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search*. MIT Press, Cambridge MA, second (2000) edition.

Tillman, R., Danks, D., and Glymour, C. (2008). Integrating locally learned causal structures with overlapping variables. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L., editors, *Advances in neural information processing systems*, volume 21, pages 1665–1672, La Jolla, CA. The NIPS Foundation.

Tillman, R. and Spirtes, P. (2011). Learning equivalence classes of acyclic models with latent and selection variables from multiple datasets with overlapping variables. In Gordon, G., Dunson, D., and Dudík., M., editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 3–15, Fort Lauderdale, FL. Journal of Machine Learning Research 15. http://jmlr.csail.mit.edu/proceedings/papers/v15/.

Vandamme, D., Fitzmaurice, W., Kholodenko, B., and Kolch, W. (2013). Systems medicine: helping us understand the complexity of disease. *Quarterly Journal of Medicine*, 106(10):891–895.

Wilde, M. and Williamson, J. (2016). Models in medicine. In Solomon, M., Simon, J., and Kincaid, H., editors, *Routledge Companion to Philosophy of Medicine*, pages 271–284. Routledge, New York and London.

Williamson, J. (2005a). *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press, Oxford.

Williamson, J. (2005b). Objective Bayesian nets. In Artemov, S., Barringer, H., d'Avila Garcez, A. S., Lamb, L. C., and Woods, J., editors, *We Will Show Them! Essays in Honour of Dov Gabbay*, volume 2, pages 713–730. College Publications, London.

Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford University Press, Oxford.

Williamson, J. (2013a). How can causal explanations explain? *Erkenntnis*, 78:257–275.

Williamson, J. (2013b). Why frequentists and Bayesians need each other. *Erkenntnis*, 78(2):293–318.

Williamson, J. (2017). *Lectures on inductive logic*. Oxford University Press, Oxford.