Citation: Pothos, E. M., Lewandowsky, S., Basieva, I., Barque-Duran, A., Tapper, K. & Khrennikov, A. (2021). Information overload for (bounded) rational agents. Proceedings of the Royal Society B: Biological Sciences, 288(1944), 20202957. doi: 10.1098/rspb.2020.2957

Information overload for (bounded) rational agents

Emmanuel M. Pothos[1]

Stephan Lewandowsky[2]

Irina Basieva[1]

Albert Barque-Duran[3]

Katy Tapper[1]

Andrei Khrennikov[4]

Affiliations and email:

1. Department of Psychology, City, University of London, London EC1V 0HB, UK. Irina.Basieva@city.ac.uk, Albert.Barque-Duran@city.ac.uk, katy.tapper.1@city.ac.uk, Emmanuel.Pothos.1@city.ac.uk
2. School of Psychological Science, University of Bristol, Bristol BS8 1TU, UK. Stephan.Lewandowsky@bristol.ac.uk
3. Department of Computer Science, Universitat de Lleida, Carrer de Jaume II, 67, 25001 Lleida, España.
4. International Center for Mathematical Modeling In Physics and Cognitive Science, Linnaeus University, Universitetplatsen 1 351 95 Växjö, Sweden. andrei.khrennikov@lnu.se

Running head: dysfunctional disagreement

Main text word count, including references: 7291

Abstract

Bayesian inference offers an optimal means of processing environmental information and so an advantage in natural selection. We consider the apparent, recent trend in increasing dysfunctional disagreement in e.g. political debate. This is puzzling because Bayesian inference benefits from powerful convergence theorems, precluding dysfunctional disagreement. Information overload is a plausible factor limiting the applicability of full Bayesian inference, but what is the link with dysfunctional disagreement? Individuals striving to be Bayesian-rational, but challenged by information overload, might simplify by using Bayesian Networks or the separation of questions into knowledge partitions, the latter formalized with quantum probability theory. We demonstrate the massive simplification afforded by either approach, but also show how they contribute to dysfunctional disagreement.

Keywords: Bayesian inference, disagreement, entrenchment, rationality, decision making

1. Background

"Truthiness is tearing apart our country ... It used to be, everyone was entitled to their own opinion, but not their own facts. But that's not the case anymore."

–Stephen Colbert, January 2006

Living organisms depend on the optimal processing of environmental information, for example, regarding foraging, mate selection, or the assessment of predation risks. Environmental information is typically uncertain, and so has to be processed probabilistically. The established standard for probabilistic inference is Bayesian Probability Theory ([1]; we will refer to it as just Bayesian theory or occasionally full Bayesian theory, for emphasis). Bayesian theory provides a set of mutually coherent principles for probabilistic reasoning on uncertain premises. Bayesian theory benefits from powerful normative arguments, such as the Dutch Book Theorem, which shows that Bayesian probabilities will never lead to inconsistencies, such as certain loss in a combination of gambles [1]. Accordingly, Bayesian reasoning is often characterized as rational. There is an immense body of work successfully validating Bayesian models of human cognition [2-4]; these models are not universally successful, but they are successful enough to allow confidence that humans can be sometimes rational in the Bayesian sense.

Moreover, for non-human animals, it has been argued that Bayesian inference confers a natural selection advantage [5-6] and there have been simulations of how natural selection enables the computation of Bayesian priors across generations [7] or other aspects of Bayesian behaviour [8] (the first step in probabilistic inference is the determination of priors, that is, the assumptions regarding the probabilities of relevant events prior to any new information). Evidence for animal behaviour consistent with Bayesian inference has been observed in, for example, foraging [9] or mating ([10]; overview in [11). The requirement of optimality in animal behaviour is often grounded in Bayesian terms, even acknowledging that Bayesian consistency may be focused on particular environments or circumstances [8,12].

However, for both humans and non-human animals, there have been inconsistencies between Bayesian principles and behaviour. For humans, some evocative examples have been produced by the influential work of Tversky and Kahneman. For example, Tversky and Kahneman described a hypothetical person, Linda, as outgoing, concerned with equality, and intellectually restless [13]. Naïve participants considered it more likely that Linda is a bank teller and a feminist, than just a bank teller. Such conjunction fallacies challenge Bayesian intuition at a fundamental level; it is like judging that it is more likely to rain and snow in December, than just snow. Interestingly, analogous fallacies appear in animal behaviour too. For example, rhesus macaques can show ambiguity aversion [14] and pigeons sometimes show the less is more effect, whereby a desirable food plus a less desirable food is perceived less appealing than the desirable food alone [15].

As Valone ([11], p.257) noted, “Greater attention needs to be devoted to understanding when and when not to expect Bayesian updating and to determine the limits of Bayesian updating in animals.” The exact point applies to human behaviour too. Here we pursue a novel perspective to the emergence of non-Bayesian behaviour in humans, motivated by the apparent increase in dysfunctional disagreement in, e.g., modern political debate. We call dysfunctional disagreement when it appears impossible for two parties to converge, regardless of iterations and evidence. Our analysis is not restricted to political debate, but it is easier to develop the argument this way.

The evidence for increasing dysfunctional disagreement and deterioration in the quality of political debate is strong. For example, consider: the emergence of “truthiness”, as in Colbert’s quote above (based on his satirical show), which can be defined as "truth that comes from the gut, not books" [16]; the increasing dissemination of “fake news” [17] and their ability to set the political agenda [18]; the intense polarization surrounding recent political events (e.g., the Brexit referendum vote in the UK). Kahan ([19], p.1) offers an evocative quote: “Never have human societies known so much about mitigating the dangers they face but agreed so little about what they collectively know.”

It is tempting to consider these points unsurprising, because there is a staggering range of factors contributing to disagreement, particularly when people rely on false information [20]. Disagreement may arise due to emotional influences. Emotion can overwhelm objective information [21] or bias the activated information [22]. Some theorists suggest that all reasoning is motivated [23], so that discourse is guided just by insistence on a particular position. Differences in values can result in persistent disagreement [24]. For example, conflicts between a refutation message for a prior position and valued self-conceptions may lead people to become more entrenched [25]. There are several related biases. For example, the disconfirmation bias is scepticism for premises incongruent with one’s beliefs [26]. The “mybias” is collecting information and assessing evidence in a way biased in favour of a person’s beliefs [27]. Mybias is especially problematic in information-rich societies, since plurality and freedom of expression mean that one can find supporting opinions for any position. For example, Del Vicario et al. [28] argued that information related to distinct narratives generates homogeneous, polarized communities on Facebook. Such echo chambers could embody contradictory perspectives between them [29] and lead to distorted pictures regarding consensus.

We focus on individuals striving to be as Bayesian as possible, be up to date with the relevant information, and be willing to put aside their egos in the interest of resolving disagreements constructively. We call such individuals *well-meaning*, and also suggest that they can set aside unmovable personal values (i.e., we need not worry about disagreement from values, [24]). Such well-meaning individuals should be able to avoid most of the 'standard' sources of disagreement. For example, in dual decision routes analytic vs. intuitive components [30] correspond to thoughtful vs. spontaneous cognition. Bayesian inference might be predominantly localized in the analytic route; but, the relative balance between different routes is partly under conscious control, depending on effort, time etc. Or Bayesian inference might be reflected in the intuitive route, with non-Bayesian behaviour arising from limitations from working memory or language when accessing the basis of intuitive judgments [31]. But, it should be possible to reduce such limitations, with effort. Also, decision biases might be avoidable with the adoption of behavioural rules [32]; it is known that emotions can be monitored and their impact on behaviour limited [33]; etc.

Here is the paradox: more people are educated than ever before in history, there is more insight regarding decision biases, we have better understanding of the importance of the common good, and access to information has never been easier. All these factors should increase our capacity for Bayesian cognition. At the very least, we can assume that the proportion of well-meaning individuals in society has not changed, maybe even increased (would we not like to consider ourselves as well-meaning?). So, why does it appear that increasingly there is dysfunctional disagreement surrounding many current debates?

We suggest that, even for well-meaning individuals, information overload challenges our capacity for Bayesian thought, in a way that leads to dysfunctional disagreement. It is easiest to make our case in relation to political debate, but the ideas are general. First, we ask whether there is increasing information overload in political debate. The case is straightforward. One cause of information overload is the multiplicity of media and ways to disseminate information in modern society. Practically every second, the internet, television , mobile phones etc. pump out massive amounts of news, comments on the news, and comments on the comments. Another cause is that, in a technologically advanced society, some debates are complex, for example, because they relate to technological innovations that cannot be easily comprehended in lay terms. Access to information has never been easier and we enjoy unprecedented benefits from technological advancement, yet these factors contribute to massive information overload.

Second, we consider whether information overload might contribute to dysfunctional disagreement. There are indications that this is the case [34]. Allenby and Sarewitz [35] suggest that the technological complexity of modern society is such that informed decisions are beyond the scope of comprehension for the majority of us. John [36] suggests that scientists best serve society by *relaxing* the maxims of transparency and openness—not because openness and transparency are undesirable, but because too much information may damage public trust in science, because the public's folk philosophy of science is at odds with the actual workings of science. There is clearly a pessimistic view concerning whether people can deal with the information complexity in modern political debates [37-38].

We develop a precise link between information overload and non-Bayesian inference and consider the implications for dysfunctional disagreement, even for well-meaning individuals. It is interesting that animal behaviour researchers have also considered whether information overload (environmental complexity) might challenge Bayesian processes [39].

2. Outline of Methods

We consider two well-meaning individuals, Alice and Bob, debating a question and examine their capacity for avoiding dysfunctional disagreement, under conditions of information overload. Convergence means agreement on at least the probabilities for question outcomes, noting that in complex debates it is rarely the case there are uncontested observations, even for good faith actors. We quantify information overload in terms of the number of ancillary questions, which inform our decision on a key question. For example, suppose Alice is interested in the Brexit question. She could inform her eventual decision on Brexit by considering questions such as 'Will Brexit be good for the economy?', 'Will Brexit be good for employment rights' etc., noting that each of these questions could be further broken down. There is information overload when the number of these ancillary questions increases beyond a 'practical' point.

Can well-meaning individuals agree to disagree? Bounded rationality is the form of rationality which emerges when the resources of the reasoning agent are insufficient for full rationality. So, what are forms of bounded rationality under conditions of information overload and the implications for dysfunctional disagreement?

3. Disagreement and Bayesian rationality

Consider well-meaning Alice and Bob debating a complex political question and assume they share their questions and outcomes. They then use their respective information to define a probability distribution and update their beliefs as rational Bayesian agents. Is it possible for Alice and Bob to dysfunctionally disagree? Suppose Alice and Bob have different information regarding a Brexit question, but share priors and have common knowledge of each other's posteriors (posteriors are the updated probabilities, once some new information has been received). Then Aumann's [40] theorem guarantees that Alice and Bob's posteriors will be the same, that is, two rational agents will eventually converge. Moreover, this convergence can be achieved with a reasonable amount of effort [41]. The requirement of common priors may appear stringent; however, it can be replaced by milder ones [42]. Even without common priors, Bayesian Alice and Bob willing to share information must eventually converge. The Bernstein, von Mises's theorem guarantees that Bayesian updating will converge posteriors (as long as there is no 'zero priors' trap, [43]). Finally, some of these results depend on honest exchange of information. For well-meaning Alice and Bob this should be straightforward, assuming they can agree on acceptable error bounds. Overall, well-meaning Bayesian Alice and Bob committed to full Bayesian inference cannot agree to disagree [41-42].

How practical is it for Alice and Bob to be fully Bayesian under conditions of information overload? The essential idea is this (see also Supplementary Material 1). Consider a finite set Ω of all possible elementary events (the most specific events which can occur) and all possible subsets, including the null set $\emptyset$ and Ω itself. This set theoretic representation of events is appropriate if each event is either true or not true[1]. We can perform logical operations on these subsets, union, intersection, and complementation, which correspond to the familiar operations of conjunction, disjunction, and negation. The requirement that each of these operations produces a subset of Ω enables an algebra over the space of subsets, which is a Boolean algebra (because the operations obey commutativity, associativity, and distributivity). We can then define a probability measure over these subsets, which is, a map from the space of subsets to the real number interval $[0,1]$, with normalization 1 for Ω.

Consider Alice confronted with questions *A*,*B*,*C*,*D*…, each of which can have possible outcomes $A_1 \ldots A_n$, $B_1 \ldots B_m$ etc. Each block of question outcomes generates its own Boolean algebra, $\beta(A)$, $\beta(B)$, … Before Alice can engage with probabilistic reasoning for a question, she first needs to construct these individual Boolean algebras, which involves a process of specifying conjunctions, disjunctions, and negations of outcomes. But, for a Bayesian Alice confronted with questions, *A*, *B*, … *F*, it is insufficient to have $\beta(A)$, $\beta(B) \ldots \beta(F)$. For a consistent *joint probability distribution* across any combination of question outcomes, she also needs to construct a bigger Boolean algebra $\beta(A, B, \ldots F)$, which integrates the algebras for the individual questions in a consistent way. This larger algebra requires knowledge of conjunctions and disjunctions for all the individual question outcomes $A_i, \ldots, F_j$, belonging to the different algebras $\beta(A)$, $\beta(B)$, … $\beta(F)$.

The problem of intractability of full Bayesian representations is well known, cf. the idea of magic sets in Artificial Intelligence [44]. We illustrate it in the case of debating e.g. Brexit and

[1] On any Boolean algebra, it is possible to define a truth function taking values 'true' or 'not true'. On a partial Boolean algebra, see below, such a truth function cannot be introduced.

ancillary questions, such as whether Brexit might be good for the economy, labour laws, etc. If we had nine binary ancillary questions, then the elementary events would be enumerated as

1. $Brexit_{yes}$, $X1_{yes}$…$X9_{yes}$
2. $Brexit_{yes}$, $X1_{yes}$…$X9_{no}$

…

1024. $Brexit_{no}$, $X1_{no}$…$X9_{no}$

Given these 2^{10}=1024 elementary events, we can evaluate any more elaborate question, for example, a conjunction involving some question outcomes vs. others, such as $Prob(X1_{yes}\&X2_{yes}\ or\ X3_{yes}\&X5_{no})$. But, the immense expressive power of Bayesian theory comes with the price of requiring knowledge of the joint probability distribution – here, the probabilities of all 1024 elementary events. The more questions we have, the more complex the joint probability distribution and so any probabilistic inference. As the number of questions *n* and outcomes per question *k* increase, the number of terms in the joint probability distribution increase as k^n.

To quantify complexity, we adopt an information-theoretic coding scheme and compute *information costs* ([45-46]; Supplementary Material 1). The coding cost of *D* numbers can be specified by dividing the relevant number range into *D* bins and assigning each number to one bin, which requires $\log_2 D$ bits for each number for a total of $D \log_2 D$ bits. This is intuitive because if the *D* numbers were uniformly distributed, we would have enough bins to just make them discriminable (if *D*=100, these statements are equivalent to representing the numbers with two decimal places; Supplementary Material 2). Therefore, the information cost for representing probabilistic information for *n* questions with *k* outcomes each is $(k^n-1)\log_2(k^n-1)$ bits, approximated as $k^n \log_2 k^n$.

Information overload clearly undermines full Bayesian inference. Consider a person living in an isolated community a hundred years ago. He would be confronted with a fairly limited range of questions, each of which would be affected by relatively few events. So, it would be undemanding to create a Boolean algebra of all questions, including conjunctions, disjunctions etc. Today, especially in political debate, we are confronted with questions of immense complexity. Consider Alice faced with the Brexit dilemma. There are hundreds of questions relevant to resolving the dilemma, across several categories, for example, relating to finance, immigration, security, and so on. Alice does not have the time or resources (mental or otherwise) to create a full Boolean algebra for all questions and their outcomes.

When confronted with a complex probability distribution, a powerful approach is sampling algorithms, such as Markov Chain Monte Carlo (MCMC) methods [3,47-48]. An MCMC method will approximate Bayesian computations, by employing samples from the probability distribution, instead of the full distribution. Such samples are often selected to favour more probable parts of the distribution and depending on the similarity of the parts already selected. However, in the present case, sampling approximations will not help: when faced with problems of increasing complexity, sampling from the full distribution will delay, but not avoid, the exponential explosion of probability terms.

4. Bayesian Networks

The first approach we consider for mitigating the problems of complex distributions is Bayesian Networks [e.g., 49]. Suppose we recognize that in many cases questions will be independent of each other, so that e.g. $Prob(A|B) = Prob(A)$ or conditionally independent so that e.g. $Prob(A\&B|X) = Prob(A|X) \cdot Prob(B|X)$. Clearly, such an approach has simplifying potential, since a complex conditional probability $Prob(A|X_1, X_2, X_3, X_4 \ldots)$ might be easily computable as e.g. $Prob(A|X_1)$. The way to formalize assumptions about conditional independence is Bayesian Networks. Bayesian Networks represent (acyclic) probabilistic relations between a set of variables, such that each variable is a node and causal relations are represented as directed edges. The simplifying potential of Bayesian Networks rests with their Markov property: without causal dependencies there are no conditional dependencies. So, simplification depends on the causal structure. Note, there is extensive evidence for the psychological plausibility of Bayesian Networks [50-51], even if it is unclear whether they suffice for a cognitive theory of causality [52]. Presently, we are only concerned with the way the local Markov property can simplify probabilistic information.

If Alice and Bob are overwhelmed by the complexity of their representations, they could use Bayesian Networks as a simplifying tactic. But it is unlikely they will develop similar causal structures for their representations, as these would depend on their experience, education, background etc. Bayesian Networks Alice and Bob with different causal structures means that the powerful classical convergence theorems (Aumann's theorem; the Bernstein, von Mises's theorem) no longer hold. Alice and Bob could now find themselves in a state of dysfunctional disagreement, even though they are fully rational given their representations (which correspond to different assumptions regarding causal structure). Alice and Bob could seek convergence by communicating their causal structure, but such knowledge is often hard to articulate. Note, there have been attempts to explain dysfunctional disagreement with Bayesian Networks with hidden nodes corresponding to e.g.

attitudes which prevent convergence [53-54]. The present point is related, but instead concerns the inevitable incidental differences in causal structures.

To estimate the complexity of probabilistic inference with Bayesian Networks, consider classical Alice contemplating six binary questions related to the Brexit question. Without the Markov property the probability distribution for a particular combination of question outcomes would look like

$$Prob(X1_{yes}, X2_{yes}, X3_{yes}, Y1_{yes}, Y2_{yes}, Y3_{yes}, Brexit_{yes}) =$$
$$Prob\left(X1_{yes} \middle| X2_{yes}, X3_{yes}, Y1_{yes}, Y2_{yes}, Y3_{yes}, Brexit_{yes}\right) \cdot$$
$$Prob\left(X2_{yes} \middle| X3_{yes}, Y1_{yes}, Y2_{yes}, Y3_{yes}, Brexit_{yes}\right) \ldots Prob(Brexit_{yes})$$

. The Markov property allows us to assume certain questions to be independent. For example, regarding $Prob(A \mid X, Y)$ we may be able to write $Prob(A \mid X, Y) = Prob(A \mid X)$. Suppose that Alice employing a Bayesian Network assumes partial conditional independence, so that conditionalizations depend on *m* variables. Then, we would write, if *m*=2,

$$Prob(X1_{yes}, X2_{yes}, X3_{yes}, Y1_{yes}, Y2_{yes}, Y3_{yes}, Brexit_{yes}) =$$
$$Prob\left(X1_{yes} \middle| A_{yes}, B_{yes}\right) \cdot Prob\left(X2_{yes} \middle| C_{yes}, D_{yes}\right) \ldots$$

, where *A*, *B* are two questions on which *X*1 depends etc. As long as *m*<<*n*, each term requires k^m probabilities (ignoring '-1'), for a total of approximately $n \cdot k^m$ probabilities [55]. The associated coding complexity for the joint probability distribution given a particular Bayesian Network is $n \cdot k^m \log_2(n \cdot k^m)$ bits. We also need the information cost of specifying a Bayesian Network, and can show that overall the information cost for probabilistic information encoded using a Bayesian Network is $(n \cdot k^m) \log_2(n \cdot k^m) + n\left[\log_2 \binom{n-1}{m} + \log_2 n\right]$ (Supplementary Material 2).

5a. Quantum Probability Theory – disagreement

We call quantum theory the probability rules from quantum mechanics, without the physics. Behaviours that appear classically erroneous can sometimes have simple explanations in quantum theory, which motivates the psychological plausibility of such models [56-58].

Informally, quantum theory is just like Bayesian theory for subsets of questions (compatible sets, see below), but across these subsets apparent classical errors can arise. These incompatible sets are like knowledge partitions, segments of knowledge such that within each segment, but not across segments, reasoning is rational. Knowledge partitions can emerge as a simplifying strategy in complex problems [59-60]. For example, when learning an association between two variables based on a complex function, a natural approach is to learn the association in smaller ranges, but in a way that the corresponding parts are not integrated with each other. Well-meaning Alice dealing with Brexit might try to be rational for specific subsets of questions, but without trying to integrate the

Boolean algebra for one theme with another. For example, if Alice works in the financial sector, she may be able to create a full Boolean structure regarding the financial implications from Brexit and so be rational for such questions. At the same time, Alice is so busy with the construction of this finance Boolean algebra, that she does not have time to do the same for other Brexit questions, e.g., relating to security. Arguably, this is what we are seeing in modern society: individuals highly knowledgeable and rational in specific areas but who, when asked to consider questions across other areas, may be challenged and even produce inconsistent beliefs.

In quantum theory, instead of a set Ω of elementary events, we have a Hilbert space *H*, such that each vector in *H* corresponds to an elementary event (a Hilbert space is essentially a complex vector space with a scalar product). Question outcomes correspond to subspaces in *H*; each subspace is associated with a projector *P* (which 'lays' down a vector onto a subspace); in psychological theory, the mental state is represented by a normalised vector in *H*; probabilities are computed by projecting the state vector onto subspaces and squaring the length of the projections. Different partitions in *H* are defined by sets of basis vectors. For example, in a standard coordinate space, we might have three basis vectors along the *x*,*y*,*z* directions. Basis sets are not unique. If we apply the same rotation to each of our current vectors *x*,*y*,*z*, we will end up with a new set of basis vectors *x'*,*y'*,*z'*. Two sets of basis vectors can be related to each other using a generalised kind of rotation.

Projectors can be compatible, in which case we have a Boolean algebra exactly as in the classical case, or incompatible, when the Boolean algebra structure breaks down. That is, considering sets A, B, C … of projectors, such that within each set projectors are compatible, but across incompatible sets, one *cannot* combine Boolean algebras $\beta(A)$, $\beta(B)$…into one large Boolean algebra. Each event in this larger structure is no longer either true or not true (before measurement) and distributivity is no longer obeyed. Instead, we have a partial Boolean algebra, which is a collection of Boolean algebras pasted together, so that where any two Boolean algebras overlap, their operations agree. Conjunctions and disjunctions preserve their Boolean features only within the same Boolean algebra. Conjunctions of incompatible questions have a sequential form and $Prob(P_A \wedge then\ P_B) \neq Prob(P_B \wedge then\ P_A)$. Also, a definite answer for a question can create uncertainty for other incompatible ones.

Quantum theory can simplify probabilistic inference with incompatibility, which allows Alice to squeeze information about, say, 100 questions (which, even if binary, will require a classical space of 2^{100} dimensions) into a space of, say, 10 dimensions. If quantum Alice organizes her large set of Brexit questions into incompatible themes, each theme corresponds to a basis set in the same small dimensionality space and the representation of new themes need only involve a change of basis, instead of enlargement of the original space. However, incompatibility contributes to dysfunctional disagreement.

One implication of incompatibility is that quantum Alice is more likely to display (classical) fallacies, which may undermine her arguments. Incompatibility has been linked with conjunction and disjunction fallacies [61], question order effects [62], violations of normative constraints in causal reasoning [51], and disjunction effects [63]. Moreover, incompatibility leads to contextuality in meaning. If quantum Alice and Bob have different partial Boolean algebras, they may think they are talking about the same question, have the same data, and fail to agree, because they are talking about different questions (Figure 1). Such ideas resemble proposals in social psychology about how earlier questions can activate thoughts or perspectives for later ones [64]. Contextuality arises in quantum theory because the meaning of question *A* is determined by considering the set of questions compatible with *A* (and some of these questions might be incompatible with each other) and because the meaning of question *A* may be affected by considering prior questions incompatible with *A*.

Contextuality contributes to dysfunctional disagreement. First, quantum Alice and Bob are no longer aided by Aumann's theorem [65]. Common knowledge in the quantum case is not equivalent to common knowledge in the classical case, because the former lacks conjunctions. Additionally, questions incompatible with common knowledge will produce interference terms so that Alice and Bob will not update probabilities consistently with each other. Second, collective decision-making typically benefits from communal knowledge effects, such as the community of knowledge effect, wisdom-of-the-crowds, and Condorcet's Jury theorem. Such effects are not specific to Bayesian inference, but they are consistent with it. However, all three are undermined by contextuality. Regarding community of knowledge, Sloman and Fernbach [66] argued that in a complex world we increasingly benefit from each other's expertise and sometimes, as a result, overestimate our own knowledge (a knowledge illusion). The wisdom-of-the-crowds effect is the proposal that an averaged judgment across observers can be more accurate than most individual judgments, assuming primarily independence of observations and that individual estimates are normally distributed around the correct outcome [67]. Finally, the Condorcet Jury theorem shows that a majority decision (e.g., in a jury) is increasingly likely to be correct, as we add voters whose (individual) probability that they are correct is just over 0.5. Regarding community of knowledge and wisdom of the crowds, if Alice and Bob are debating contextual question *A*, then Alice may be thinking of A_X and Bob of A_Y, where *X*, *Y* indicate differing meanings. This casts doubt on the rationality of putting Alice's and Bob's intuitions together. Such problems are likely to be accentuated, because employing a partial Boolean algebra may lead to overconfidence (Supplementary Material 3).

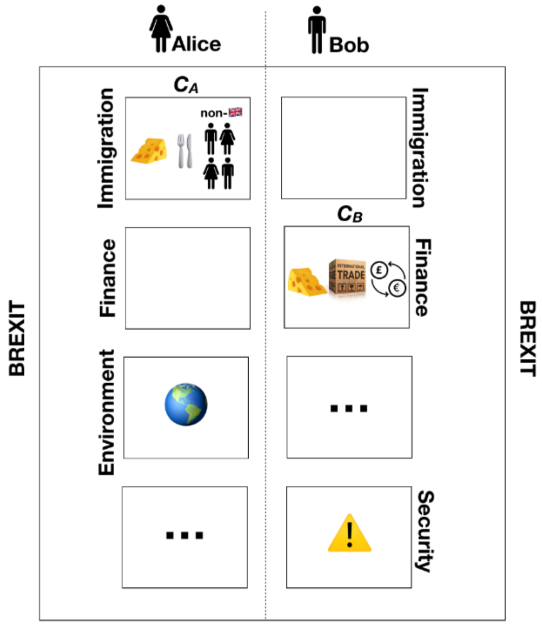


Figure 1. Alice and Bob are interested in whether Brexit may increase the price of imported cheese, *C*. Alice considers *C* with questions related to immigration, while Bob with finance questions. As a result, Alice and Bob develop meanings for the *C* question which are different, even though they think they are considering the same question.

5b. Quantum Theory – coding costs

Within a single partition, we have a classical probability distribution for the corresponding questions, encoded in the mental state vector. We need to specify the mental state for one partition and the way partitions relate to each other; the latter is encoded in transformation operators called unitary. So, the information cost for probabilistic inference for quantum Alice depends on three elements, the mental state vector for one partition, unitary operators, and the cost of allocating questions to partitions. The mental state vector and unitary operators are specified in terms of parameters which are real numbers. Regarding information costs, we follow from the above approach to assume that *F* real parameters (assumed in a certain range) can be approximately specified using $F \log_2 F$ bits.

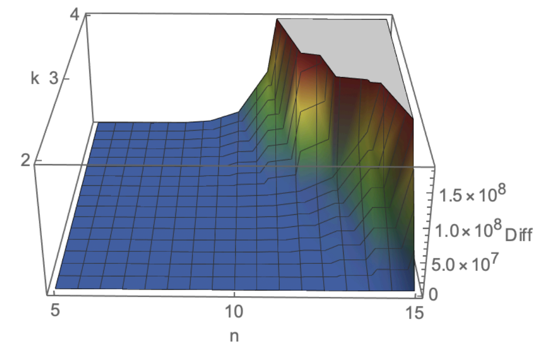
Label the dimensionality of each partition as *N*. The mental state vector in *N* dimensions has *N*-1 real parameters corresponding to amplitudes and *N*-1 real parameters for the phases. This is because the *N* amplitudes are constrained by the normalization condition and, regarding the *N* phases, the quantum state is the same up to an overall phase factor. The corresponding information cost is $2 \cdot (N-1)\log_2(N-1)$, which can be approximated as $2 \cdot N \log_2 N$. What is *N*? Suppose *c* partitions are employed and that all partitions have the same number of questions. Then, in each partition we have n/c questions, *k* outcomes each, so that $N = k^{n/c}$. The overall information cost involves additional terms, for how information in one partition relates to information in other partitions. This cost is $2 \cdot k^{n/c} \log_2 k^{n/c} + (c-1) \cdot \frac{4n}{c} \log_2 \frac{4n}{c} + \log_2 \frac{n!}{\left[\left(\frac{n}{c}\right)!\right]^{c-1}} + \log_2 \frac{(c-1)! n^c}{c^{c-1}}$ (Supplementary Material 2). Note, the dimensionality of quantum Alice's probability space turns out to be only $N = k^{n/c}$, which seems like a huge saving compared to Bayesian Alice for whom $N = k^n$; but this simplification is partly offset by the complexity of specifying partition relations.

6. Comparisons

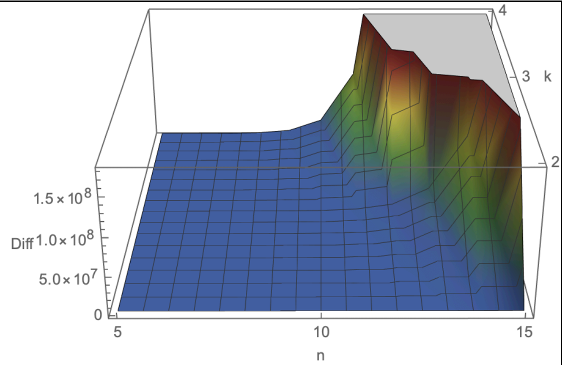
A well-meaning Alice overwhelmed by the complexity of her joint probability distribution might seek to simplify the representations either by employing Bayesian Networks or dividing her questions into (incompatible) partitions. For the latter two schemes, the critical parameters are, respectively, *m* (the average number of questions each one question depends on) and *c* (the number of partitions). Both parameters concern the extent of dependence of questions amongst themselves and, specifically, the length of conditional probabilities (Supplementary Material 2). Regarding *m*, this interpretation follows directly from the definition of a Bayesian Network, while in the quantum case classical conditionalization occurs only within knowledge partitions. Therefore, it is natural to set $\frac{n}{c} = m$ or $c = \frac{n}{m}$.

We provide indicative estimates regarding the simplification from Bayesian Networks and quantum theory relative to Bayesian theory, varying question numbers from 5 to 15 and question outcomes from 2 to 4, Figure 2. The vertical axis shows information cost for scheme A (e.g., Bayesian theory) minus B (e.g., Bayesian Networks). Recall, lower information costs are more advantageous, so that when A-B>>0, then B is superior to A. In all cases, probabilistic reasoning with either Bayesian Networks or quantum theory affords overwhelming simplification relative to Bayesian theory. This is a demonstration of the essential point that information overload will drive even well-meaning Alice to make representational approximations, putatively employing Bayesian Networks or knowledge partitions.

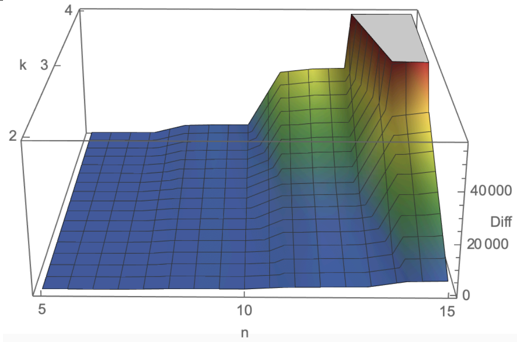
We also observe a marginal advantage of quantum theory over Bayesian Networks, though this conclusion is sensitive to the complexity of the relation between partitions. Overall, the quantum approach to simplification seems advantageous, thus providing a strong expectation of dysfunctional disagreement due to incompatibility and partitions.



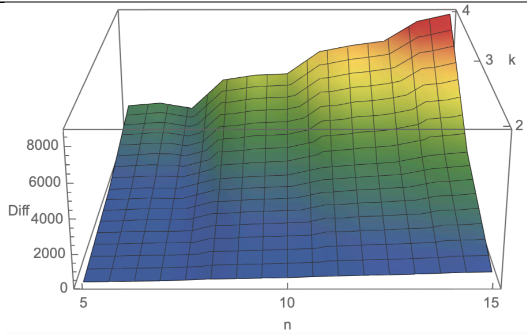
Bayesian theory minus Bayesian Networks (m=n/3).



Bayesian theory minus quantum theory (c=3).



Bayesian Networks (m=n/3) minus quantum theory (c=3).



Bayesian Networks (m=3) minus quantum theory (c=n/3).

Figure 2. We plot information cost given one scheme minus information cost given another scheme, labelled Diff (in bits). The superior scheme has lower information cost. Horizontal axes represent number of questions (n) and outcomes per question (k); complexity increases with both n and k. Note, m=3 for Bayesian Networks translates to three questions per knowledge partition in QPT.

7. Concluding comments

We considered how dysfunctional disagreement can arise for well-meaning individuals, because of information overload. The notion of being well-meaning is primarily underwritten by an assumption of rational cognition, in the Bayesian sense. There is a strong consensus that Bayesian rationality is achievable *to some* extent [1-4]. Our aim has been to understand how information overload can challenge full Bayesian rationality, how Bayesian Networks and quantum theory offer flavours of limited or local Bayesian rationality, and the implications for dysfunctional disagreement.

Regarding dysfunctional disagreement, a full Bayesian would quickly find it impossible to build the required Boolean algebra, for complex problems. Alice can simplify with Bayesian

Networks, truncating her probability distributions with assumptions about the causal structure between her questions. Alice and Bob may find themselves failing to converge if their Bayesian networks are different; Aumann's [40] and the Bernstein, von Mises's theorems no longer hold. Alternatively, Alice can simplify using knowledge partitions [59] dividing her questions into sets, such that within each knowledge partition she is fully Bayesian, but across partitions apparent errors arise. With knowledge partitions, Aumann's and the Bernstein, von Mises's theorems also no longer hold and, in addition, the resulting contextuality challenges the community of knowledge effect [66], wisdom-of-the-crowds [67], and Condorcet's Jury theorem.

Is it possible for Bayesian Networks or quantum Alice and Bob to converge? In the former case, they need to share their causal structure. However, it seems unlikely this would occur, because we are often unaware of the causal dependencies impacting on inference. In the latter case, Alice and Bob need to share their partitions (and information on how partitions relate to each other), and in addition be careful to respond to a question in the same context (Figure 1). We agree with Lissack [68] who argued that truthiness can be reduced if Alice and Bob "Try to see things from my viewpoint." However, we think quantum Alice and Bob will not engage with such a process, because contextuality is not recognized in probabilistic inference.

Our focus has been dysfunctional disagreement, because this is an under-researched topic and because the link with information overload is intuitive. More generally, there have been long research traditions concerning the way complexity undermines Bayesian rationality. The present framework can shed light into other instances of behaviour apparently problematic from a full Bayesian perspective, because of complexity, bearing in mind that there will be behaviours outside any probabilistic framework. For example, the emergence of some conjunction fallacies, as in the Linda example [13], could be traced to lack of familiarity with partition combinations. It is possible that we have a local partition for professions and one for personal characteristics, like feminism, without making the effort to combine them together. Conversely, the less is more effect in animal behaviour [15] seems harder to understand as complexity-driven bounded rationality.

In closing, to the long list of factors contributing to dysfunctional disagreement, we add differences in causal structure and contextuality, from information overload. A surprising implication is that more information or nuanced perspectives may exacerbate disagreement by further encouraging truncated probability distributions or incompatible representations as simplifying tactics. For some important modern debates, such as Brexit, it may seem that we have forgotten how to evaluate arguments using easily verifiable facts, but increasing information may not help or indeed be harmful [36-38]. A precise understanding of the impact of information overload, as we have offered, will hopefully contribute to mitigating interventions.

Acknowledgements

This work was supported by the Office of Naval Research Global [grant number N62909-19-1-2000]. Thanks to Adam Sanborn for MCMC advice and three anonymous reviewers. The original idea linking information overflow and incompatibility was AK's. Computer code in https://osf.io/h9tjk/?view_only=0041f73a47fa40cd83b89cb5a8a53e54.

References

1. Oaksford, M. & Chater, N. (2009). Précis of Bayesian rationality: the probabilistic approach to human reasoning. Behavioral and Brain Sciences, 32, 69-120.
2. Griffiths, T.L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J.B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences, 14, 357-364.
3. Sanborn, A.N., Griffiths, T.L., & Navarro, D.J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. Psychological Review, 117, 1144-1167.
4. Tenenbaum, J.B, Kemp, C., Griffiths, T.L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. Science, 331, 1279-1285.
5. Oaten, A. (1977). Optimal foraging in patches: a case for stochasticity. Theor. Popul. Biol., 12, 263-285.
6. Iwasa, Y, Higashi, M. and Yamamura, N. (1981). Prey distribution as a factor determining the choice of optimal foraging strategy. American Naturalist, 117, 710-723.
7. Ramirez, J. C. & Marshall, J. A. R. (2017). Can natural selection encode Bayesian priors? Journal of Theoretical Biology, 426, 57-66.
8. McNamara, J.M., Green, R.F., & Olsson, O. (2006). Bayes' theorem and its application in animal behaviour. Oikos, 112, 243-251.
9. Valone, T. J. & Giraldeau, L.A. (1993). Patch estimation by group foragers: what information is used? Animal Behavior, 45, 721-728.
10. Luttbeg, B. and Warner, R.R. (1999). Reproductive decision-making by female peacock wrasses: flexible versus fixed behavioral rules in a variable environment. Behavioral Ecology, 10, 666-674.
11. Valone, T. J. (2006). Are animals capable of Bayesian updating? An empirical review. Oikos, 112, 252-259.

12. Trimmer, P. C., Houston, A. I., Marshall, J. A. R., Mendl, M. T., Paul, E. S., & McNamara, J. M. (2011). Decision-making under uncertainty: biases and Bayesians. Animal Cognition, 14, 465-476.
13. Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjuctive fallacy in probability judgment. Psychological Review, 90, 293-315.
14. Hayden, B.Y., Heilbronner, S.R., & Platt, M.L. (2010) Ambiguity aversion in rhesus macaques. Frontiers in Neuroscience, 4, 1–7
15. Zentall, T. R., Laude, J. R., Case, J. P., & Daniels, C. W. (2014). Less means more for pigeons but not always. Psychonomic Bulletin & Review, 21, 1623-1628.
16. Meyer, D. (2006). The Truth of Truthiness. CBS News.
17. Allcott, H. & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31, 211-36.
18. Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media & Society, 20(5), 2028-2049.
19. Kahan, D.M. (2015). What is the "science of science communication"? Journal of Science Communication, 14, 1–2.
20. Lewandowsky, S., Ecker, U.K.H., Seifert, C.M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. Psychological Science in the Public Interest, 13, 106-131.
21. Wilson, R.S. & Arvai, J.L. (2006). When less is more: how affect influences preferences when comparing low and high risk options. Journal of Risk Research, 9, 165-178.
22. Bower, G.H. (1981). Mood and memory. American Psychologist, 36, 129-148.
23. Kunda, Z. (1990). The Case for Motivated Reasoning. Psychological Bulletin 108, 480–498.
24. Schwartz, S.H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), Advances in experimental social psychology (Vol. 25, pp. 1–65). New York, NY: Academic Press.
25. Trevors, G.J., Muis, K.R., Pekrun, R., Sinatra, G.M., & Winne, P.H. (2016). Identity and epistemic emotions during knowledge revision: a potential account for the backfire effect. Discourse Processes, 53, 339-370.
26. Taber, C.S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. American Journal of Political Science, 50, 755-769.
27. Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. Current Directions in Psychological Science, 22, 259-264.

28. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., & Quattrociocchi, W. (2016) The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113, 554-559.
29. Pariser, E. (2011). The filter bubble: what the internet is hiding from you. Penguin: London, UK.
30. Kahneman, D. (2001). Thinking fast and slow. Penguin: London, UK.
31. Oaksford, M. & Hall, S. (2016). On the source of human irrationality. Trends in Cognitive Sciences, 20, 336-344.
32. Gollwitzer, P.M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology, 38, 69-119.
33. Chambers, R., Gullone, E., & Allen, N. B. (2009). Mindful emotion regulation: An integrative review. Clinical Psychology Review, 29(6), 560-572.
34. Lissack, M. (2016). Don't be addicted: The oft-overlooked dangers of simplification. Journal of Design Economics and Innovation, 2, 29-45.
35. Allenby, B.R. & Sarewitz, D. (2011). We've made a world we cannot control. New Scientist, 14th May, 28-29.
36. John, S. (2018). Epistemic trust and the ethics of science communication: against transparency, openness, sincerity and honesty. Social Epistemology, 32, 75-87.
37. Hills, T.T. (2019). The dark side of information proliferation. Perspectives on Psychological Science, 14, 323-330.
38. Sunstein, C.R. (2020). Too much information: Understanding what you don't want to know. MIT Press: Cambridge, MA.
39. Valone, T. J. (1992). Information for patch assessment: a field investigation with black-chinned hummingbirds. Behavioral Ecology, 3, 211-222.
40. Aumann, R.J. (1976). Agreeing to disagree. Annual Statistics, 4, 1236-1239.
41. Aaronson, S. (2005). The complexity of agreement. Proceedings of ACM STOC, 634–643.
42. Hanson, R. (2006). Uncommon priors require origin disputes. Theory and Decision, 61, 319–328.
43. Basieva, I., Pothos, E.M., Trueblood, J. Khrennikov, A., & Busemeyer, J.R. (2017). Quantum probability updating from zero priors (by-passing Cromwell's rule). Journal of Mathematical Psychology, 77, 58-69.
44. Halevy, A. (2007). Dataspaces: A new paradigm for data integration. In Brazilian Symposium on Databases.

45. Oaksford, M. & Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection. Psychological Review, 101, 608-631.
46. Pothos, E. M. & Chater, N. (2002). A simplicity principle in unsupervised human categorization. Cognitive Science, 26, 303-343.
47. Lloyd, K., Sanborn, A., Leslie, D., & Lewandowsky, S. (2019). Why higher working memory capacity may help you learn: sampling, search, and degrees of approximation. Cognitive Science, e12805.
48. Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. Cognitive Science, 38(4), 599–637.
49. Gershman, S.J., Vul, E., & Tenenbaum, J.B. (2012). Multistability and perceptual inference. Neural Computation, 24, 1-24.
50. Rehder, B. (2014). Independence and dependence in human causal reasoning. Cognitive Psychology, 72, 54–107.
51. Trueblood, J.S., Yearsley, J.M., & Pothos, E.M. (2017). A quantum probability framework for human probabilistic inference. Journal of Experimental Psychology: General, 146, 1307-1341.
52. Goldvarg, E. & Johnson-Laird, P.M. (2001). Naïve causality: a mental model theory of causal meaning and reasoning. Cognitive Science, 25, 565-610.
53. Jern, A., Chang, K.M.K., & Kemp, C. (2014). Belief polarization is not always irrational. Psychological Review, 121, 206-224.
54. Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. Topics in Cognitive Science, 8, 160–179.
55. Murphy, K. (1998). A brief introduction to graphical models and Bayesian Networks. https://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html
56. Busemeyer, J.R. & Bruza, P. (2011). Quantum models of cognition and decision making. Cambridge University Press: Cambridge, UK.
57. Haven, E. & Khrennikov, A. (2013). Quantum Social Science. Cambridge University Press: Cambridge, UK.
58. Pothos, E.M. & Busemeyer, J.R. (2013). Can quantum probability provide a new direction for cognitive modeling? Behavioral & Brain Sciences, 36, 255-327.
59. Lewandowsky, S., Kalish, M., & Ngang, S.K. (2002). Simplified learning in complex situations: knowledge partitioning in function learning. Journal of Experimental Psychology: General, 131, 163-193.

60. Kalish, M.L., Lewandowsky, S., & Kruschke, J.K. (2004). Population of linear experts: knowledge partitioning and function learning. Psychological Review, 111, 1072-1099.
61. Busemeyer, J.R., Pothos, E.M. & Franco, R., Trueblood, J.S. (2011) A quantum theoretical explanation for probability judgment ‘errors’. Psychological Review, 118, 193-218.
62. Wang, Z., Solloway, T., Shiffrin, R.M., & Busemeyer, J.R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. PNAS, 111, 9431-9436.
63. Pothos, E.M. & Busemeyer, J.R. (2009). A quantum probability explanation for violations of 'rational' decision theory. Proceedings of the Royal Society B, 276, 2171-2178.
64. Schwarz, N. (2007). Attitude construction: Evaluation in context. Social Cognition, 25, 638-656.
65. Khrennikov, A. & Basieva, I. (2014). Possibility to agree on disagree from quantum information and decision making. Journal of Mathematical Psychology, 1–15.
66. Sloman, S. & Fernbach, P. (2017). The knowledge illusion: the myth of individual thought and the power of collective wisdom. Macmillan: London, UK.
67. Surowiecki, J. (2005). The Wisdom of Crowds. Anchor Books: New York.
68. Lissack, M. (2017). Explain the mess we are in: How did we get Trump, Brexit, dissonance, uncertainty and worse? WOSC17 Congress in Rome, “Science with and for Society”, Talk January 29, 2017. https://www.youtube.com/watch?v=Wqb2WYy_-Rw

Supplementary Material 1 – dynamics and additional details for quantum theory.

Dynamics for Bayesian theory. In the main text we focused the complexity discussion for both Bayesian theory and quantum theory on static representations. However, decision models are invariably dynamic, so that probabilistic decision models invoke the Bayesian and quantum rules for how probabilities change with time. The complexity picture is essentially unchanged; we offer in this subsection some corresponding notes for Bayesian theory and in the next subsection for quantum theory, also including some additional details for quantum theory,

The most basic mechanism for probabilistic updating in Bayesian theory is Bayesian updating, based on Bayes's law. But Bayesian theory also allows for a dynamical evolution of probabilities. Generally, for each question *A*, *B*, *C*… $Prob_A = (Prob_A(x,t), x = 1, \ldots k)$ is a probability vector with the index *x* enumerating question outcomes. For each $Prob_A(t)$, the system of Kolmogorov forward equations is $\frac{dProb_A(t)}{dt} = K_A \cdot Prob_A(t)$, where K_A, the intensity matrix for question *A*, is a transition matrix which determines which elements of $Prob_A(x,t)$ grow more or less probable. For a collection of questions, we write $\frac{dProb(i,t)}{dt} = \sum_{j=1}^{k^n} K_{ij} \cdot Prob(j,t)$, where the index *i* selects a term in the complete joint distribution, and the summation over *j* is over all other terms (marginals need not be enumerated separately as they are recoverable from the joint). The Bayesian dynamical picture is dynamics on a family of vectors; however, for any realistic situation with $n, k \gg 1$ we would have a large number of differential equations.

Quantum theory is likely to be unfamiliar to many readers, from either animal or human behaviour. Nevertheless, it is essentially just a way to assign probabilities to question outcomes, alternative to Bayesian theory. Quantum theory is an important part of the quantum mechanics theory of physics, but it can be employed in any situation where there is a need to quantify uncertainty. We offer here Figure SM1, which helps illustrate some of the basic ideas in quantum theory. Recall that question outcomes are subspaces. In the same way we can have a set of basis vectors for the entire space, we can also define a subspace with a set of basis vectors. In Figure SM1, we consider three question outcomes. E, ~E (not E), and G, all represented as one-dimensional subspaces.

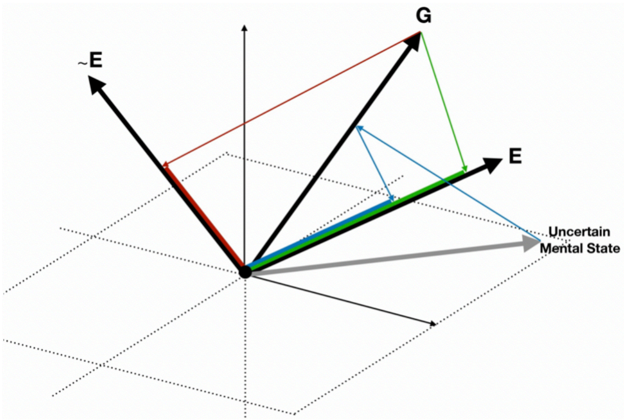


Figure SM1. We consider two incompatible binary questions, (G, ~G; we only show the former) and (E, ~E). Recall, probabilities are computed as the squared length of projections. First, consider a mental state along the G ray (a ray is a one-dimensional subspace). This means that the decision maker is certain about this question outcome, G. However, this certainty implies unavoidable uncertainty for the E question, since there are non-zero projections from G to the E (red) and ~E (green) rays. Therefore, it is impossible to resolve both questions concurrently. Second, consider an uncertain mental state (as labelled). From such a mental state resolving the G and then E question (shown) will produce a different probability than resolving first E and then G (not shown). This illustrates the non-commutativity of projectors for incompatible questions.

Regarding the dynamical evolution of probabilities in quantum theory, the analogue of the forward Kolmogorov equation in quantum theory is Schrödinger's equation, which is $\frac{d\psi(t)}{dt} = -iH\psi(t)$, so that $\psi(t) = e^{-i \cdot t \cdot H}\, \psi(0) = U(t)\psi(0)$, where H is the Hamiltonian, a transition matrix which determines which elements of $\psi(t)$ increase or decrease in amplitude, and $U(t)$ is a unitary operator. If we have compatible questions, then the dimensionality of the space and dynamics are equivalent to the Bayesian case. For example, for independent questions *A*, *B* , the overall Hamiltonian can be written as a sum of tensor products, $H = H_A \otimes I_B + I_A \otimes H_B$, so that $\frac{d\psi(t)}{dt} = e^{-i \cdot t \cdot (H_A \otimes I_B + I_A \otimes H_B)} = e^{-i \cdot t \cdot H_A} \otimes e^{-i \cdot t \cdot H_B}\psi(t)$, where the state vector matches the structure of the Hamiltonian, in the expanded space. For incompatible questions, there is one Hamiltonian for all questions and the time evolved state can be used to answer any question. For example, $\frac{d\psi(t)}{dt} = -iH\psi(t)$, $Prob\big(A; \psi(t)\big) = |P_A\psi(t)|^2$ and for another question B (which may not even be known in advance), $Prob\big(B; \psi(t)\big) = |P_B\psi(t)|^2$, where P_A, P_B can be related by a unitary transformation. Outcome combinations for incompatible questions also do not evolve separately, e.g., $Prob\big(A \wedge thenB;\ \psi(t)\big) = |P_B P_A \psi(t)|^2$.

Note, in Bayesian theory the dynamical equation operates directly on probabilities, so given a Bayesian initial state obeying the law of total probability, any time-evolved state will also obey the law of total probability. By contrast, in quantum theory the dynamical equation operates on amplitudes, which lead to probabilities using Born's rule. So, an initial state can be made to obey the law of total probability, but a time-evolved state need not do so (Pothos & Busemeyer, 2009).

We can now consider the complexity situation for the Bayesian and quantum dynamical evolution of probabilities. Essentially, the comparative picture for relative complexities does not change, but a detailed complexity calculation will be unnecessarily involved. For completeness, we offer some brief notes.

It is straightforward to see that incompatibility simplifies not only representation, but also dynamical processing. Recall, if Bayesian Alice considers questions $A_1, A_2 \ldots A_n$, each with *k* outcomes, then she needs one differential equation for each term in the joint probability distribution $Prob(A_1 = A_{1_i}, A_2 = A_{2_j} \ldots A_n = A_{n_z})$ and then the probability for a particular outcome for a question is recovered from marginalizing across these (k^n) terms. The increase in equations is exponential in number of questions and *n*-power in question outcomes. Quantum Alice considering incompatible questions $A_1, A_2 \ldots A_n$ needs a single differential equation to determine the outcome of a single question A_1 ($2 \cdot k$ real equations), regardless of *n*. If quantum Alice introduces more questions, then for each one of them she needs to determine a unitary transformation relating the new question basis to a canonical one, whose specification requires maximally $\sim k \times k$ equations (in practice, we would expect quantum Alice to employ far fewer constraints in determining the unitary transformation). Relatedly, quantum Alice is able to encode more efficiently (some) interrelatedness information in dynamical processing, relative to a Bayesian Alice. Consider a single question, *k* outcomes. Both Bayesian and quantum Alice need to specify their mental state, *k* vs. (approximately) 2*k* values, assuming quantum Alice employs a superposition. Bayesian Alice can include interrelatedness information in the intensity matrix, but new interrelatedness patterns require different intensity matrices (size $\sim k \times k$). If quantum Alice's Hamiltonian is incompatible with the question operator (this would be generally the case for non-trivial dynamics), then interrelatedness information in phase differences will impact on time-evolved amplitudes.

Additional references

Pothos, E.M. & Busemeyer, J.R. (2009). A quantum probability explanation for violations of 'rational' decision theory. Proceedings of the Royal Society B, 276, 2171-2178.

Supplementary Material 2 – coding costs supplements

In this section we provide some additional technical details, regarding coding costs for full Bayesian Alice, Bayesian Networks Alice, and quantum Alice.

First, we consider the basic problem of encoding *D* probabilities. At the heart of the argument in main text is the way complexity of probabilistic inference is quantified. The most basic problem is how to represent *D* probabilities. Note, this representation must be approximate, since otherwise we are left with real numbers and the information cost of specifying a real number is infinite. The initial proposal is that if we have *D* probabilities, then minimally we need to employ *D* bins in the relevant range, so that each probability is in principle discriminable assuming they are uniformly distributed. In practice, this assumption will rarely be true, so that if we insist on complete discriminability we may need more bins in certain parts of the range and fewer bins in other parts of the range; for *D* numbers, *D* can be considered a reasonable, on average estimate for the number of bins required.

We next consider whether the constraint that probabilities have to sum to 1 can reduce the information cost for approximately representing *D* probabilities. We first have to order the probabilities from largest to smallest, which costs $\log_2 D!$ bits. Regarding the assignment of the first, largest probability, we have *D* possibilities. Regarding the assignment of the second probability, note that the first and the second largest probabilities cannot sum to greater than 1. Therefore, for the second largest probability, the available choices are *D*/2 at most. For example, if the first probability is higher than 0.5, then the second has to be lower than 0.5, hence the number of available bins would be fewer than *D*/2. Alternatively, if the first highest probability is lower than 0.5, then both the first and the second highest probabilities have to be lower than 0.5; in either case, we (still) have fewer than *D*/2 available bins for assigning the second probability. When assigning the third largest probability, likewise the available choices are *D*/3 etc., until the smallest probability. So, overall, the total number of possible assignments is given by

$$[D]\cdot\left[\frac{D}{2}\right]\cdot\left[\frac{D}{3}\right]\cdot\ldots\left[\frac{D}{D}\right]=\frac{D^D}{D!}$$

The corresponding information cost is $\log_2\frac{D^D}{D!}$, so the total (taking into account the cost of ordering probabilities too) is $\log_2\frac{D^D}{D!}+\log_2 D!=\log_2 D^D$, as before. Therefore, the normalization constraint for numbers which are probabilities cannot reduce the information representation cost, compared to assuming we just have numbers in a certain range.

There are several alternative coding schemes regarding probabilities specifically. For example, suppose Alice initially places all probabilities in the first bin. Then, she considers how many of these probabilities would be high enough to be assigned to (at least) the second bin. These probabilities are at most D/2, so we need to select $D/2$ items out of D – there are $\binom{D}{D/2}$ ways of doing so, requiring $\log_2 \binom{D}{D/2}$ bits. This procedure can be repeated until we run out of bins. However, in each step we also need to specify the number of probability terms which go forward (to the next bin), requiring $\log_2 D$. Another coding scheme would involve again starting with ordering probabilities. Then Alice knows that after assigning the *largest* number the next one cannot be larger than $1 - 1/D$, the second largest one cannot be larger than $1 - 2/D$ etc. This is because the first probability is the highest one and the lowest value for this probability is $1/D$. The smaller number of bins that Alice can drop for consideration after the first assignment is one. So, for the first probability she has *D* bins available, for the second number *D*-1 bins available etc. This means that the information cost of assigning all probabilities to bins (i.e., representing all probabilities) is $2 \log_2 D!$ However, a simple computational analysis shows that such alternative schemes are generally inferior to the proposed one, that *D* probabilities require $D \log_2 D$ bits for their approximate representation.

There are two more issues to consider. First, there have been proposals for adaptive approaches for estimating probabilities, based on the observed frequencies. However, such proposals do not concern the representation of (just) the numbers corresponding to the various probabilities. That is, presently, we are not interested in estimating a probability from observed frequencies, rather the cost of representing the numbers corresponding to the different probabilities. Second, it might be tempting to employ the actual probabilities to specify an entropy-like code. Recall the definition of Shannon's entropy, which is that for objects x_1, x_2, x_3…with probabilities p_1, p_2, p_3, the most efficient code per object is given (on average) by its entropy measure. However, the code for each object is different from the code required to represent the probability – a number. Put differently, if an object is more likely, its Shannon code will be lower because the frequency of the object will be higher, but there is no sense in which a probability number p_1=0.02 will need more or fewer bits than p_2=0.98, since in both cases we are representing a number with a required precision.

Second, we consider some additional detail concerning the information cost of specifying the structure of a Bayesian Network. A Bayesian Network has a number of nodes equal to the number of questions, *n*. For each node, we have a fan-in of *m*. We need to identify which *m* connections, out of a possible *n*-1 ones, are made to this particular node, and there are $\binom{n-1}{m}$ ways to select *m* elements from *n*-1 ones. This requires $\log_2 \binom{n-1}{m} + \log_2 n$ bits for each node (note, the information cost is unchanged for each node because there are no restrictions in the number of times a particular node can connect to other nodes). The final cost for the structure is given by $n\left[\log_2\binom{n-1}{m}+\log_2 n\right]$, because we have *n* different nodes and also we need to encode the cost for specifying the integer *m*, which is $\log_2 n$ (this term will typically be dwarfed by the rest).

Third, we consider some additional detail concerning the information cost of representing probability information with quantum theory. Recall, in main text we noted how in quantum theory question outcomes correspond to subspaces. As noted, a subspace is specified by a set of basis vectors, which is a collection of orthonormal vectors, called eigenstates, that span the subspace. A partition in the overall space can also be defined by a set of basis vectors. The idea of basis vectors is essential in understanding how partitions can be related to each other, with unitary transformations.

Each basis vector for a partition corresponds to a unique combination of outcomes in the partition. For example, suppose we have a partition with three binary compatible questions X1, X2, X3, then the partition will be eight-dimensional (2x2x2). Each of the eight basis vectors will have the form $X1_{yes}X2_{yes}X3_{yes}$, $X1_{yes}X2_{yes}X3_{no}$, etc. Consider a different partition of Y1, Y2, Y3 binary questions, compatible with each other and incompatible with the *X* ones. A unitary operator relates basis vectors in one partition to basis vectors in the other, which means how a particular combination of outcomes for *X* questions depends on particular combination of outcomes for *Y* questions. In the most complex case, a particular combination of question outcomes for the *X* questions can depend on *all possible* combinations of question outcomes for the *Y* questions (in N dimensions, the corresponding unitary would have N^2 parameters). We think that psychologically this is implausible.

It is straightforward to show that knowledge of a sequential conjunction for two incompatible questions allows one constraint in the specification of the corresponding unitary operator. The more the conjunctions that quantum Alice can specify between the *X* and *Y* question outcomes, the richer the eventual specification of *U* (that is, the richer Alice's understanding of the relation between the two knowledge partitions). Note, this discussion shows that there is potentially more structure in a quantum theory representation than in a Bayesian Networks one (cf. Pothos et al., 2017). The role of Bayesian Networks is to simplify dependences between variables, but a Bayesian Network itself does not provide guidance regarding how one conditional probability should relate to another. By contrast, in quantum theory the separation of questions into knowledge partitions needs to be accompanied by information on how the questions in one partition relate to the ones in others.

How much effort will quantum Alice plausibly invest in specifying the relation between knowledge partitions? Consider Dirlam's (1972) estimate of optimal chunk sizes, assuming a hierarchically organized memory. He suggested that at each node in the hierarchy there should be three to four branches – and so three to four associations with other elements in the hierarchy. Also, limits in environmental sampling have been related to additional reinforcement when learning high correlations (Hourihan & Benjamin, 2010; Kareev, 2000) or the facilitation of complex learning through a more structured development of the relevant knowledge (Elman, 1993; Newport, 1990; Plunkett & Marchman, 1993); such limits might restrict quantum Alice's ability to develop a complex understanding of the relation between knowledge partitions. Likewise, we suggest that quantum Alice will seek to understand the relation between partitions employing only a few constraints per relation, as $4n/c$ per *U* for *n* questions and *c* partitions. Assuming there are *c* knowledge partitions, quantum Alice needs to specify the relation between any one of them and a canonical one, so that the information cost of the corresponding *U*'s is $(c-1)\cdot\frac{4n}{c}\log_2\frac{4n}{c}$ (as above, since for each *U* we have to represent $4n/c$ real numbers).

We next consider the information cost of dividing questions into *c* partitions. The dimensionality of each partition is $N = k^{n/c}$; for example, $N = 8$ indicates that we have clusters of three binary questions. Since each partition has n/c questions, we need to identify which n/c questions out of *n* ones belong to it. This is given by $\log_2\binom{n}{\frac{n}{c}} + \log_2 n$ for the first partition, $\log_2\binom{n-\frac{n}{c}}{\frac{n}{c}} + \log_2\left(n-\frac{n}{c}\right)$ for the second partition etc., for a total of $\sum_{i=0}^{c-1}\left[\log_2\binom{n-i\frac{n}{c}}{\frac{n}{c}} + \log_2\left(n-i\frac{n}{c}\right)\right] = \sum_{i=0}^{c-1}\left[\log_2\frac{\left(n-i\frac{n}{c}\right)!}{\left(\frac{n}{c}\right)!\left(n-(i+1)\frac{n}{c}\right)!} + \log_2\left(n-i\frac{n}{c}\right)\right]$. Regarding the first term in the summation, observe that the numerator for $i = c-1$ is part of the denominator for $i = c-2$, and so on, so all these terms simplify to give $\log_2\frac{n!}{\left[\left(\frac{n}{c}\right)!\right]^c}$. Also, the second term in the summation amounts to $\log_2\frac{(c-1)!n^c}{c^{c-1}}$.

Overall, as stated in main text, if quantum Alice considers *n* questions with *k* outcomes each, divided into *c* equally sized partitions, the information cost is $2\cdot k^{n/c}\log_2 k^{n/c} + (c-1)\cdot\frac{4n}{c}\log_2\frac{4n}{c} + \log_2\frac{n!}{\left[\left(\frac{n}{c}\right)!\right]^{c-1}} + \log_2\frac{(c-1)!n^c}{c^{c-1}}$, arranged so that we consider first the cost of the mental state, then the cost for the unitaries capturing the relations between partitions, and finally the cost of allocating questions to partitions.

<u>Additional references</u>

Dirlam, D. K. (1972). Most Efficient Chunk Sizes, Cognitive Psychology, 3, 355-359.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small, Cognition, 48, 71-99.

Hourihan, K. L. & Benjamin, A. S. (2010). Smaller is better (when sampling from the crowd within): low memory-span individuals benefit more from multiple opportunities for estimation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 1068-1074.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. Psychological Review, 397-402.

Newport, E. L. (1990). Maturational constraints on language learning. Cognitive Science, 14, 11-28.

Plunkett, K. & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. Cognition, 48, 21-69.

Pothos, E. M., Busemeyer, J. R., Shiffrin, R. M., & Yearsley, J. M. (2017). The rational status of quantum cognition. Journal of Experimental Psychology: General, 146, 968-987.

Supplementary Material 3

We think that employing a partial Boolean algebra is likely to lead to overconfidence. A psychological sense of uncertainty is often quantified using entropy, $S = -\sum_i p_i \log p_i$, where the summation ranges across all outcomes to a question. Entropy is higher when more options are equiprobable and the more certain we are regarding the resolution of a question, the lower the entropy; e.g., if regarding a binary question we have *Prob*(Yes)=0.9, entropy will be lower compared to if *Prob*(Yes)=0.6. Consider quantum Alice contemplating the Brexit issue, which consists of several specific questions. The entropy function is additive and so Alice's total entropy will be the sum of individual question entropies. Suppose Alice simplifies her Brexit Boolean algebra, so that she considers only 2-3 questions in her preferred basis set. Given the small number of questions, she can plausibly devote sufficient effort to each question and move from a state of higher uncertainty to one of lower uncertainty (e.g., with binary questions, suppose that initially *Prob*(Q1, yes)=0.6, *Prob*(Q2, yes)=0.4, *Prob*(Q3, yes)=0.5, but after some thought *Prob*(Q1, yes)=0,8, *Prob*(Q2, yes)=0.1, and *Prob*(Q3, yes)=0.9).

Suppose Bob employs a more faithful Boolean algebra, consisting of 20 questions. Bob will have a more accurate, nuanced picture for Brexit. However, if we assume that Alice and Bob have the same amount of time for their deliberation, then Bob will be able to devote less time per question than Alice, and so the reduction in uncertainty for Bob's (already more numerous) questions will be lower than that for Alice. After deliberation, on average, Alice is likely to end up with questions of lower entropy than Bob (Figure SM2). Additionally, the maximum possible entropy increases with the dimensionality *N* as $N \log N$. So, if information overflow encourages Alice to squeeze a complicated dilemma into a small space (using incompatibility), Alice may end up being more confident than Bob, even though her representation is less accurate. There is some indirect support for this idea. First, it appears that increasing information can increase confidence, without increasing accuracy (e.g., Chervany and Dickson, 1974; Davis et al., 1994; Paese and Sniezek, 1991). Second, the Dunning-Kruger effect is the observation that low ability individuals can have a harder time recognizing their limitations and so are more likely to feel overconfident (Kruger & Dunning, 1999).

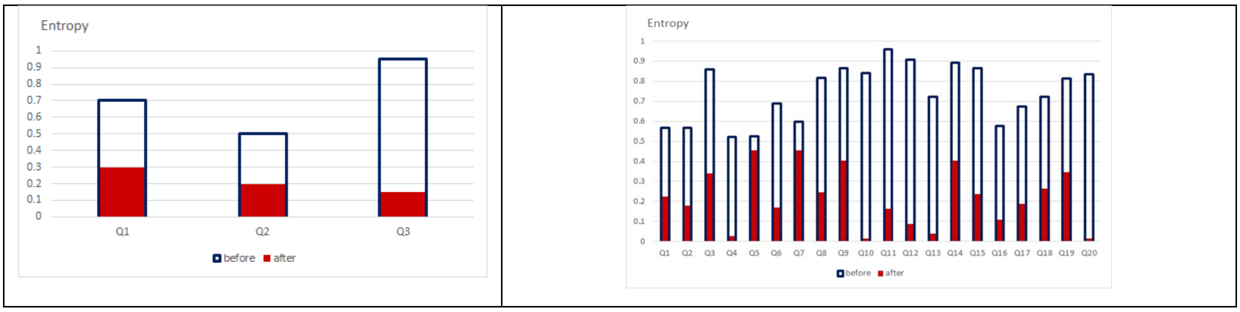


Figure SM2. Alice employs a more simplified representation for her problem and so can devote more time per question than Bob, assuming that Alice and Bob have the same amount of time for their deliberation. The blue outline shows uncertainty before deliberation and the red filler after deliberation. Alice may end up resolving to a more satisfactory extent her fewer questions – and so feel more confident than Bob – but this is largely because Bob's picture was more nuanced and accurate to start with.

Supplementary Material 4—some additional computational results

For Bayesian Networks, more costly representations will involve a more complex causal structure, when *m*=*n*/3; this is the version we compared with full Bayesian theory in main text. We show here that even more simplification can be achieved if the causal structure is simpler, with *m*=3 (Figure SM3a). For quantum theory, more costly representations will involve fewer, larger partitions when c=3; this is the version we compared with Bayesian theory in main text. We show here that greater simplification can be achieved when there are more numerous, simpler partitions, with *c*=*n*/3.

Overall, Bayesian Networks will afford more simplification when *m*=3 than when $m = n/3$ (conditionalizations are simpler in the former case). Quantum theory will afford more simplification when $c = n/3$ than when *c*=3 (there are more partitions in the former case). So, the versions of Bayesian Networks and quantum theory considered here are even more advantageous relative to Bayesian theory, compared to the versions in main text.

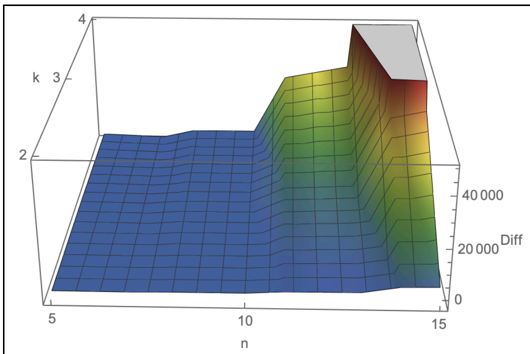


Figure SM3a. Bayesian Networks (*m*=*n*/3) minus Bayesian Networks (*m*=3). The graph shows that the more complex causal structure (*m*=*n*/3) is most costly than the simpler one (*m*=3).

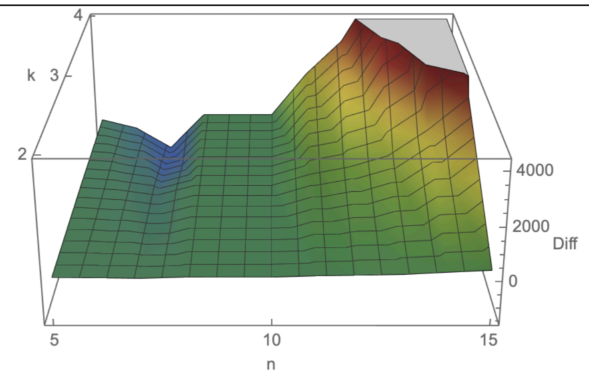


Figure SM3b. Quantum theory (*c*=3) minus quantum theory (*c*=*n*/3). The graph shows that fewer, larger partitions (*c*=3) are more costly than more numerous ones (*c*=*n*/3).

The final issue is whether the overwhelming advantage of coding schemes based on Bayesian Networks or quantum theory, over full Bayesian theory, can be reduced, if some sampling approach is incorporated in the coding schemes. We think this is not the case. We can demonstrate this by offering variants of the top two panels in Figure 2 in main text, but with an assumption that only 0.01% of the probability terms comprising the full distributions are encoded (we do this conservatively and approximately, by reducing the probability terms, but not scaling down any of the other costs). Observing Figures SM4a and SM4b, it is clear that an exponential increase in complexity, with increasing questions and question outcomes still occurs. So, our essential point (that Bayesian Alice will be challenged by the information cost of complex debates) remains, even if there are two mitigating factors concerning the urgency of simplification: first, reducing the probability terms means that there will be many situations for which full Bayesian will be as good as or even better than an approach based on Bayesian Networks or quantum theory. This is evident in the figures, because we are plotting only data points for which full Bayesian encoding is inferior. Where the figures show blank, full Bayesian encoding is superior (e.g., when *n*=10 and we are considering binary questions). Second, the onset of the exponential increase in complexity occurs later. So, Bayesian Alice invoking sampling approximations will be confounded by information overload only after more questions and outcomes per questions, compared to Bayesian Alice without sampling approximations. Notice that in main text Figure 2 the vertical axis for 'Diff' (the information cost advantage) extends to 1.5×10^8, whereas presently this extends to only about 40,000, given the same ranges for *n*, *k*.

Notwithstanding these points, please also bear in mind that we have explored the impact from a massive reduction in probability terms – 0.01% reduction means that for 10 binary questions, instead of considering 1024 probability terms to represent her probability information, sampling Alice will only consider less than one term (let's say one term). Clearly, in such cases we have to consider just how much accuracy Alice is willing to sacrifice.

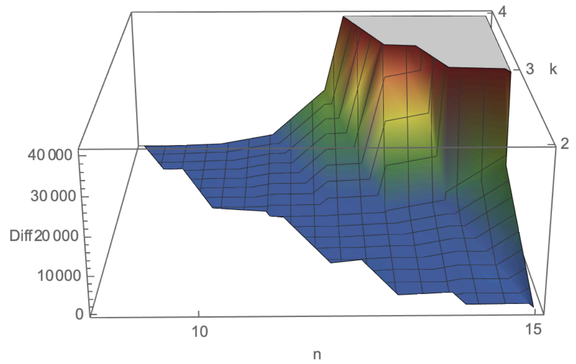


Figure SM4a. Bayesian theory minus Bayesian Networks (*m*=*n*/3), retaining only 0.01% of probability terms. Points for which Diff<0 are not plotted.

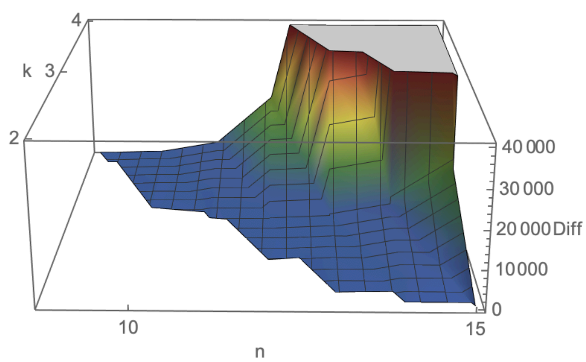


Figure SM4b. Bayesian theory minus quantum theory (*c*=3), retaining only 0.01% of probability terms. Points for which Diff<0 are not plotted.