Bridging the gap between subjective probability and probability judgments: the Quantum Sequential Sampler

Jiaqi Huang* and Jerome R. Busemeyer*
Indiana University

Zo Ebelt and Emmanuel M. Pothos
City, University of London

Abstract

One of the most important challenges in decision theory has been how to reconcile the normative expectations from Bayesian theory with the apparent fallacies that are common in probabilistic reasoning. Recently, Bayesian models have been driven by the insight that apparent fallacies are due to sampling errors or biases in estimating (Bayesian) probabilities. An alternative way to explain apparent fallacies is by invoking different probability rules, specifically the probability rules from quantum theory. Arguably, quantum cognitive models offer a more unified explanation for a large body of findings, problematic from a baseline classical perspective. This work addresses two major corresponding theoretical challenges: first, a framework is needed which incorporates both Bayesian and quantum influences, recognizing the fact that there is evidence for both in human behavior. Second, there is empirical evidence which goes beyond any current Bayesian and quantum model. We develop a model for probabilistic reasoning, seamlessly integrating both Bayesian and quantum models of reasoning and augmented by a sequential sampling process, which maps subjective probabilistic estimates to observable responses. Our model, called the Quantum Sequential Sampler, is compared to the currently leading Bayesian model, the Bayesian Sampler (Zhu, Sanborn, & Chater, 2020) using a new experiment, producing one of the largest datasets in probabilistic reasoning to this day. The Quantum Sequential Sampler embodies several new components, which we argue offer a more theoretically accurate approach to probabilistic reasoning. Moreover, our empirical tests revealed a new, surprising systematic overestimation of probabilities.

Keywords: Probabilistic Reasoning; Sequential Sampling; Quantum Cognition; Bayesian; Probabilistic Fallacies

One of the most theoretically important and practically significant problems in cognitive science is to understand human probabilistic reasoning. A vexing and enduring challenge has been how to reconcile an expectation of Bayesian rationality with extensive evidence of apparent paradoxes and fallacies. We will review some of the predominant Bayesian approaches to understanding fallacies and propose a new probabilistic reasoning model, based on the alternative probability rules, from quantum theory.

The development of probabilistic reasoning theory to its current state of the art, including both Bayesian variants and quantum models, is compelling: the position of Bayesian rationality is that, in probabilistic reasoning and decision making generally, human behavior ought to be consistent with the principles of Bayesian probability theory. There are powerful formal arguments as to why this should be the case (Oaksford & Chater, 2007). For example, the Dutch Book Theorem states that probability systems consistent with a particular set of requirements inoculate a reasoner from incoherent assignments of probabilities to events, that is, assignments which allow a sure loss in a betting situation (de Finetti, Machi, & Smith, 1993). Bayesian probability theory is consistent with the requirements for the Dutch Book Theorem, as well as other important results, for example, concerning the convergence of posterior probabilities (Aumann, 1976) and the practicalities of Bayesian convergence (Aaronson, 2005; Hanson, 2006). There is a large body of evidence in favor of Bayesian cognitive models, in a wide range of situations, including categorization, learning, and reasoning (e.g., Griffiths et al., 2010; Sanborn, Griffiths, & Navarro, 2010; Tenenbaum et al., 2011).

---

* The two authors are co-first authors.

All data and data analysis codes are available on Github: `https://github.com/adamhuang11111/quantum_sequential_sampler_public`.

EMP was supported by AFOSR grant FA8655-23-1-7220, and JRB was supported by AFOSR grant FA9550-20-1-0027. All authors contributed to all parts of this work.

The primary contributions were as follow: JRB oversaw all mathematical work and developed the quantum part, while JH the Markov part; JH carried out the computational work; ZE conducted the empirical investigation; EMP wrote the initial drafts of this manuscript.

Part of this work was presented here: Huang, J., Busemeyer, J. R., Ebelt, Z. & Pothos, E. M. (2023). Quantum Sequential Sampler: a dynamical model for human probability reasoning and judgments. In Proceedings of the 2023 Annual Conference of the Cognitive Science Society.

Additionally, the optimality embodied in Bayesian reasoning has been argued to confer adaptive, evolutionary advantage, for example, for foraging (Valone & Giraldeau, 1993) or mating (Luttbeg & Warner, 1999). For many non-human animals, statistical estimation of environmental information has a very tangible evolutionary value (McNamara, Green, & Olsson, 2006; Trimmer et al., 2011; Valone, 2006). If there is even a small evolutionary advantage from Bayesian processes in behavior, across generations, we expect a trend for increasing conformity with Bayesian constraints. The current evidence seems to support such views.

The picture of unquestionable benefits from Bayesian reasoning has to be moderated by the problem that full Bayesian reasoning is, in fact, intractable for any finite agent. These observations have a long history, notably with the proposal of bounded rationality (Simon, 1955), work which was recognized with a Nobel prize (for Simon in 1978). Bayesian researchers have been aware of these limitations and there has been extensive effort to develop versions of Bayesian reasoning suitable for finite agents (Gershman, Horvitz, & Tenenbaum, 2015; Griffiths, Lieder, & Goodman, 2015; Howes, Lewis, & Vera, 2009). For example, Lieder and Griffiths (2019) offered a framework for bounded Bayesian reasoning and argued that many instances of apparent deviations from Bayesian prescription can be explained as Bayesian reasoning with limited resources. As another example, in an application of Bayesian reasoning with datasets of realistic size, Lake, Salakhutdinov, and Tenenbaum (2015) employed a combination of Bayesian Networks and other simplifications to tackle the problem of learning how to recognize handwritten characters. Overall, when we encounter human behavior apparently at odds with Bayesian reasoning, it is reasonable to ask whether maybe there is an underlying Bayesian component to behavior, which through simplification or other approximations, leads to the apparent errors and fallacies.

Bayesian probability is the most established framework for probabilistic reasoning, whether in cognitive modeling or in science more generally (e.g., Howson & Urbach, 1993). Nevertheless, it is not the only one (e.g., Narens, 2014). In fact, there is an infinite hierarchy of probability systems, ordered in terms of the complexity of the basic sum rule (that is,

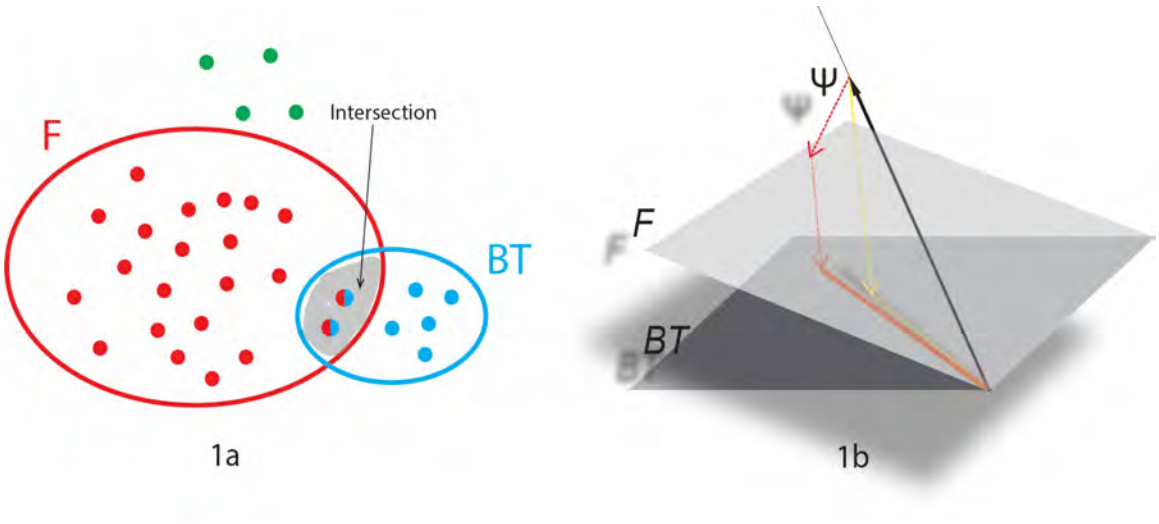


Figure 1. The two figures show Bayesian (Figure 1*a*) and quantum (Figure 1*b*) representations of the information in the Linda problem (Tversky & Kahneman, 1983). In the Bayesian case, the conjunction corresponds to the intersection between the feminist (F) and bank teller (BT) sets. In the quantum case, the conjunction corresponds to the sequential projection to the F and then BT subspaces, of the mental state vector ψ. In Figure 1b, the darker plane represents BT and the lighter plane represents F; the yellow projection represents judging BT alone and the orange projection represents judging F and then BT (both projections are along the BT subspace).

the complexity of the law of total probability; Sorkin, 1994). Below we will introduce a probability system related to Bayesian theory, but with a sum rule just a bit more complex than the Bayesian one, quantum theory. Quantum and Bayesian theories are based on different axioms and offer different predictions for how basic probabilities combine to produce more complex probabilities. Interestingly, quantum theory satisfies the Dutch Book theorem too (Pothos et al., 2017). Quantum and Bayesian theory underwrite two different hypotheses for probabilistic reasoning. There have been several proposals of successful quantum cognitive models, adding credibility to the notion that, sometimes, quantum, rather than Bayesian, principles offer a better approach to understanding probabilistic reasoning (Bruza, Wang, & Busemeyer, 2015; Busemeyer & Bruza, 2011; Haven & Khrennikov, 2013; Pothos & Busemeyer, 2013, 2022).

Another consideration is that, even though the present focus is probabilistic reasoning, the corresponding models and ideas might well turn out to be more general. Part of the appeal of probabilistic models rests in their general applicability, offering promise that

successful application in one area might translate to novel theory and prediction in other areas. For example, there have been proposals of Bayesian models in just about all areas of cognitive psychology, including learning (e.g., Griffiths & Tenenbaum, 2009; Steyvers et al., 2003), memory (Steyvers, Griffiths, & Dennis, 2006), perception (Chater, 1996), language (Griffiths & Kalish, 2007; Xu & Tenenbaum, 2007), and logical reasoning (Oaksford & Chater, 1994), as well as probabilistic reasoning. Analogously, quantum theory has been applied in models for conceptual structure (Aerts, 2009; Aerts, Sozzo, & Veloz, 2015; Bruza, Kitto, Ramm, & Sitbon, 2015), memory (Brainerd et al., 2015; Trueblood & Hemmer, 2017), similarity (Epping et al., 2023; Pothos et al., 2013), and even attentional dynamics (Atmanspacher & Filk, 2010; Rosner et al., 2022). The scope of applicability of such models is underwritten by an assumption that large parts of cognition can be understood as the processing of statistical structure.

For all their promise, probabilistic models, whether classical or quantum, do not exhaust approaches in probabilistic reasoning theory. For some researchers, instead of probability theory (Bayesian or quantum), a better route to understand probabilistic reasoning is heuristics and biases (Hertwig et al., 2013; Kahneman, Slovic, & Tversky, 1982). For example, Lopez-Astorga, Ragni and Johnson-Laird (2021) conclude, in relation to conditional probabilities, that "the probability calculus supplements human intelligence rather than underlies its deliberations". Without doubt, heuristics often embody deep intuitions about human cognition. However, heuristics and biases are sometimes imprecisely expressed or have a narrow focus. For example, Tversky and Kahneman (1983; Shafir, Smith, & Osherson, 1990) proposed to explain the conjunction fallacy with the representativeness heuristic, according to which a similarity process drives probabilistic judgments in the Linda scenario. This approach has been criticized as being vague and unsuitable for detailed empirical predictions (Moro, 2009). Additionally, using heuristics, it is hard to see how one can predict results consistent with violations of probability identities and the cancellation of noise terms in Costello and Watts (2014). As another example, the fast-and-frugal approach is quantitatively expressed and so, for example, can provide specific predictions regarding the way

naive observers answer questions like which city between Bristol and Bath has the higher population (Gigerenzer & Goldstein, 1996). However, as currently specified, this model cannot be applied to general probabilistic judgments.

Despite these shortcomings, heuristics provide important insight to probabilistic reasoning. Notably, there are aspects of behaviour beyond formal probabilistic approaches and in these cases heuristic and biases approaches come into their own. Moreover, heuristics such as representativeness and availability have descriptive value and offer alternative explanatory narratives, which complement those from formal models. That is, in some cases, probabilistic models attempt to offer an explanation combining and consistent with heuristic accounts. For example, a heuristic such as representatives tells us that probabilistic reasoning might share some elements with similarity processes – this perspective to explanation is complementary, not mutually exclusive, to that from a formal model, such as quantum theory (Busemeyer et al., 2011). Another example of this point is Trueblood et al.'s (2017) model of causal inference, which involved a quantum model, which could capture a range of heuristic accounts, as well as a Bayesian influence (see Rehder, 2014). Bayesian researchers have also tried to re-express heuristics within their frameworks (e.g., Lieder & Griffiths, 2019). While we acknowledge the importance of heuristics in explaining probabilistic reasoning, in the present work, we do not consider them in detail. Rather, our focus is on exploring the capacity of formal probabilistic models to describe probabilistic reasoning, across a large range of judgments.

A final preliminary remark is that, even if we accept the merits of approaching probabilistic reasoning theory with a formal probability framework, it seems unlikely that probabilistic principles as such would suffice for a complete explanation. A reasonable expectation is that a full model should include some process assumptions of how probabilities and/ or responses are produced. For example, even if the mind embodies probabilistic principles which are consistent with Bayesian prescription, the way relevant probabilities are estimated might be faulty or there might be a noisy response mechanism. Accordingly, apparent fallacies could be explained in a way which still allows a statement of Bayesian rationality for

humans. All these points produce interesting conundrums, regarding whether fallacies show fault with probabilistic principles versus whatever additional mechanisms the mind employs in the production of probabilities or responses. Earlier models for the conjunction fallacy and probabilistic fallacies in general have been focused on just the probabilistic framework – this includes our own model (Busemeyer et al., 2011; see also e.g. Tentori et al., 2013). More recent work has been adopting a more complete approach (Costello & Watts, 2018; Zhu, Sanborn, & Chater, 2020).

Apparent probabilistic fallacies

No finding has had as much influence in probabilistic reasoning theory than Tversky and Kahneman's (1983) conjunction fallacy. Participants were told of a hypothetical person, Linda, who was described very much like a feminist and not at all like a bank teller. They were then asked to rank order how likely different statements about Linda are. The three statements of interest concern whether she is a bank teller (BT), a feminist (F), and the conjunction between two ($F \wedge BT$). Participant ratings typically indicate that $P(F) > P(F \wedge BT)$, as would be expected, but also that $P(F \wedge BT) > P(BT)$. The latter finding challenges a fundamental principle in Bayesian theory, that a conjunction can never be more likely than a marginal. At the root cause of the problem is the fact that probabilities in Bayesian theory need to conform with set-theoretical constraints. So, in the same way that it is impossible to have more blue and red balls in an urn, than just blue balls (blue and red balls are a subset of just blue balls), it is likewise impossible to have $P(F \wedge BT) > P(BT)$, in classical probability theory.

The conjunction fallacy has proven robust across a large number of disambiguations, clarifications, and other manipulations (Dulany & Hilton, 1991; Moro, 2009; Tentori, Bonini, & Osherson, 2004). For example, researchers have considered whether a frequentist presentation of the relevant information might make participants less prone to a conjunction fallacy, since, the argument goes, probabilistic reasoning based on frequencies would be more natural for humans, than based on subjective probabilities (e.g., Gigerenzer, 1994; Sanborn &

Chater, 2016). Another suggestion has been that making the set-theoretic structure of a problem more salient can foster compliance with Bayesian theory. For example, Tentori, Bonini, and Osherson (2004) asked participants whether a Scandinavian person was more likely to have blond hair versus blond hair and blue eyes; in such a case, the relevant probabilities directly correspond to countable instances, as opposed, for example, to subjective probabilities for a single case, such as Linda – of course, the two are formally equivalent, but perhaps not subjectively so. Such manipulations can reduce the conjunction fallacy rate, but they rarely eliminate it completely (Moro, 2009).

Regarding disambiguations, there is the possibility that, perhaps, participants misunderstand the question in a way that implies there is no longer a fallacy (cf. Tentori, 2021). Dulany and Hilton (1991; Hilton, 1995; Hilton & Slugoski, 2001; see also Adler, 1984) considered how so-called conversational implicatures (Grice, 1975) might be relevant in the way participants understand the various statements in the Linda conjunction fallacy example. The proposal is that participants might be assuming that when the BT predicate is presented by itself, $\neg F$ is also implied, that is $P(BT) = P(BT \wedge \neg F)$. Of course, if the BT statement is augmented in this way, then there is no longer a fallacy, since the probability that Linda is a bank teller and a feminist can easily be higher than the probability of another conjunction. The prediction from this account is that when the question about just BT is properly disambiguated, the rate of conjunction fallacy should be greatly diminished (Dulany & Hilton, 1984; Macdonald & Gilhooly, 1990). However, there have been several studies testing this prediction and in most cases a conjunction fallacy could still be identified (e.g., Agnoli & Krantz, 1989, and Messer & Griggs, 1993, as well as the original Tversky & Kahneman, 1983, study; review in Moro, 2009).

One way to disambiguate potentially unclear statements, such as a marginal in isolation, has been to introduce a fuller range of probability judgments: for example, Tentori et al. (2004) and Wedell and Moro (2008) also included $BT \wedge \neg F$ (or the equivalent of) in conjunction fallacy experiments, so as to prevent participants from mistakenly inferring BT to be $BT \wedge \neg F$. To sum up this point, it is always possible that participants do not

understand a probability judgment as intended. The current evidence suggests that a more complete set of probability judgments will make it less likely that participants will employ unintended interpretations. Indeed, the most recent work on probabilistic fallacies, such as that from Costello and Watts (2014) and Zhu et al. (2020), has employed an increasingly more expansive set of probability judgments. Our work is in this vein.

The conjunction fallacy is nonsensical from a baseline Bayesian perspective. How can people decide that there are more Scandinavian individuals with both blond hair and blue eyes, than just blue eyes? It seems that the conjunction fallacy is such a simple result that it is tempting to imagine that it can be explained and presumably immediately corrected by anyone. However, as Stephen J. Gould famously said (1992, p.469) about the conjunction fallacy, "I know that the conjunction is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me - 'but she can't be just a bank teller; read the description' ". At the same time, Bayesian theory is intuitive too: Bayesian theory has been described as "common sense reduced to calculations" (Laplace, 1816, cited in Perfors et al., 2011). Arguably, one of the drivers of the applicability of Bayesian theory in cognition is exactly the fact that Bayesian principles are intuitive since, after all, it is human intuition that we are trying to model. The conjunction fallacy exemplifies this clash between two different, equally powerful, intuitions, and the persistence (Gilboa, 2000) of this clash has had a defining influence in the field of probabilistic reasoning and decision making.

The singular influence of the conjunction fallacy in the literature on probabilistic reasoning should not obscure the fact that there have been several other apparent fallacies, which challenge any picture of reasoning based on baseline Bayesian theory. Notably, there are disjunction fallacies, according to which people judge $P(A \vee B) < P(B)$, even though a disjunction will always be at least as likely than either of its individual premises (Carlson & Yates, 1989). There are disjunction effects, whereby people judge $P(A) \neq P(A \wedge X) + P(A \wedge \neg X)$ thus violating the classical law of total probability (Broekaert et al., 2020; Shafir & Tversky, 1992). There have also been reports of unpacking effects, when the probability of a 'packed' disjunction is judged as lower than the sum of mutually exclusive 'unpacked'

components (Tversky & Koehler, 1994). Moreover, question order effects have been observed, so that pairs of yes/ no questions are responded to differently, depending on the order in which they are presented (Moore, 2002). A final example of this non-exhaustive list of apparent fallacies is errors in the way conditionalizing information impacts on probability updating (Bergus et al., 1998; McKenzie, Lee, & Chen, 2002; Trueblood & Busemeyer, 2011) and problems with estimating conditional probabilities generally (Lopez-Astorga, Ragni, & Johnson-Laird, 2021).

Understanding human probabilistic reasoning is a challenge of proposing a framework which encompasses as many of the findings generally considered fallacies as possible. There are some ideas which, though promising for particular results, have not generalized well. We have already briefly encountered Tversky and Kahneman's (1983) representativeness heuristic. Tentori, Crupi, and Russo (2013) suggested that conjunctions are evaluated using inductive confirmation. For example, in the Linda problem, participants estimate the likelihood for the conjunction as the result of evaluating a confirmation measure that reflects the increase in probability from the initial judgment for $P(F|BT)$ to $P(F|BT \wedge$ story) when introducing the information about the story. This account works well for the conjunction fallacy, but it is difficult to adapt it to other probabilistic judgments, such as disjunctions or conditional probabilities (Busemeyer et al., 2015). As a final example, averaging accounts have been proposed for the conjunction fallacy, which purport that conjunctions are evaluated as the averages of the probabilities of each conjunct individually (Abelson, Leddo, & Gross, 1987; Fantino et al., 1997; Nilsson et al., 2009). Such accounts encounter difficulty when it comes to explanations of conditional dependencies between items, as well as the way conjunction fallacies vary depending on the causal strength between the conjuncts.

There are other examples of judgment theory focused on a single or a few apparent fallacies, but with limited capacity for generalization. The conjunction fallacy especially has attracted enormous attention, but, ultimately, this is a one degree of freedom finding (either $P(A \wedge B) < P(B)$ or $P(A \wedge B) > P(B)$) and so is poorly suited for comprehensively testing complex theories. Overall, the value of models focused on a single or a handful of effects is

unquestionable, not least because in such models it is often possible to acquire substantial insight into the reasons for good or bad performance. At the same time, there is a natural trend in the field towards more encompassing accounts.

Two recent theories have been evaluated against larger sets of probability judgments. Costello and Watts (2014) examined their theory against marginals, two-way conjunctions in different combinations of two conjuncts and their negations, and disjunctions. Even though no detailed model fits were carried out, Costello and Watts (2014) considered several probabilistic identities, purported to allow tests of their account. Costello and Watts (2016) extended the range of probabilities judgments to 10 judgments, for five pairs of weather events, so that marginals, conjunctions, disjunctions, and conditionals were included. Zhu, Sanborn, and Chater (2020) provided a more extensive empirical examination of human probabilistic judgments, by asking participants to estimate the probabilities of 20 unique questions about weather events, marginals, conjunctions, conditionals etc, involving the pair {icy, frosty}; there were another 20 questions, involving the pair {normal, typical}. Zhu et al. (2020) compared the fit of their model to this dataset and the one from Costello and Watts (2014).

Overall, there has been a trend in examining theories of probabilistic reasoning against larger datasets, which encompass all of marginals, conjunctions, disjunctions, and conditionals. From the perspective of any formal model, whether Bayesian or not, more comprehensive evaluations are essential, since the strength of such models lies exactly in how different probability terms constrain each other. For example, in a baseline Bayesian model, for three events, the three-way joint probability distribution (eight probabilities constrained to sum to one) allows the specification of all other probabilities: conjunctions, disjunctions, conditionals etc. The empirical examinations from Costello and Watts (2014, 2016) and Zhu et al. (2020) go some way towards addressing these issues, because, for a particular pair of events, several possible probability questions were considered. However, we think that the descriptive power of formal probabilistic models is better assessed against several pairs of events, assessed concurrently. In this work, we consider three events, all three event pairs,

and for each pair all conjunctions, disjunctions, and conditional probabilities - altogether 78 probabilistic judgments per participant. Such a set of probability terms subsumes individual apparent probabilistic fallacies, including conjunction fallacies, disjunction fallacies, and violations of the law of total probability. Perhaps more importantly, the inter-dependence of probabilities amongst each other would offer a more sensitive test of probabilistic models. While we think that the motivation for a new, more expansive dataset is reasonably clear, is there a need for corresponding theoretical development too?

Theoretical progress

We can summarize a large portion of recent theoretical progress in probabilistic reasoning in terms of three main ideas. The first main idea is that reasoning has a kernel of Bayesian influence, but in a way that is noisy or biased, so that deviations from strict Bayesian prescription can arise. Noise (or bias) can be motivated in several ways. For example, in Lieder and Griffiths (2019) there is a tradeoff between Bayesian consistency and resource limitations and in Dasgupta et al. (2020) probability updating can be noisy. A proposal particularly relevant to us is the one from Costello and Watts (2014), since their work offers a more complete calculus for probabilistic reasoning. In Costello and Watts' (2014) probability plus noise model, probability judgments are based on sample frequencies computed from a fixed number of samples generated from memory. For example, if a person wants to decide whether they are likely to enjoy a camping trip, they will invoke from memory previous instances of camping trips and assess the probability of enjoyment against the relevant frequencies, in a way akin to the availability heuristic from Tversky and Kahneman (1983). When there are no prior relevant instances, such as when we are called to evaluate probabilities about Linda, whom we have never encountered before (unless we are cognitive psychologists), a mental simulation process can generate such instances and employ them to compute probabilities. It can be questioned whether a mental simulation process is a plausible mechanism for generating probabilities, still, this is a common assumption in corresponding models (e.g., see Costello & Watts, 2016, p.120, or Zhu et al., 2020, p. 2).

In the work by Costello and Watts (2014), these sampling mechanisms are subject to faulty evaluations, and it is these errors which can give rise to apparent fallacies. Specifically, they define d as the probability of an evaluation error (e.g., evaluate a memory as true when in fact it was false), from which they derive the prediction $P(\text{judged } A) = (1 - 2d) \cdot P(\text{truly } A) + d$, but then they add the assumption that the error rate is higher for conjunctions so that $P(\text{judged } A \wedge B) = (1 - 2 \cdot [d + \Delta d]) \cdot P((\text{truly } A \wedge B) + [d + \Delta d]$, and analogously for disjunctions, with $\Delta d > 0$. The key point is that more complex probability evaluations, such as conjunctions and disjunctions, suffer from higher noise, relative to simpler probabilities (marginals). For both simple and complex probabilities, this model implements a regression to the mean, so that true probabilities below 0.5 increase towards 0.5 and true probabilities above 0.5 decrease towards 0.5. This model can explain the occurrence of conjunction fallacies. For example, in the Linda problem, assume that participants 'truly' judge $P(BT)$ to be only slightly greater than $P(F \wedge BT)$, such that in both cases the probabilities are less than 0.5. Regression to the mean for the conjunction is faster (because there is higher noise) than for the individual question, so that the apparent $P(F \wedge BT)$ increases more so than $P(BT)$, leading to a conjunction fallacy. A similar explanation can be used to explain apparent disjunction fallacies. Moreover, Costello and Watts (2014, 2018) examined several probabilistic identities, expected to hold under their model, and reported evidence that this is indeed so.

There are some difficulties with this proposal. For example, it cannot accommodate conjunction fallacies when the true probabilities of the marginals and conjunction are over 0.5, however, there is evidence for conjunction fallacies in that range (Yearsley & Trueblood, 2018). Also, the probability plus noise model predicts a regression towards the mean for all probabilities, even for events manifestly impossible ('there is a flying cat made of cheese outside my house') or manifestly possible ('there is water on earth'). A more theoretical concern with this model is the assumption that the sampling process for generating $P(A)$ is different from the one to generate $P(A \wedge B)$. In other words, the person makes a judgment about A using one sample and then disregards this sample, to generate a separate

sample for$A \wedge B$. This seems wasteful and indeed other Bayesian researchers have argued that "...many real-world tasks require making many decisions based on the same information. In these scenarios, it makes sense for an agent to cache and reuse samples for several decisions" (Vul et al., 2014, p.29). If the probability plus noise model employs a single sample for marginals and conjunctions, then no conjunction fallacy can emerge. A possible shortcoming of Costello and Watts' (2014) model, shared with the quantum probability one (Busemeyer et at., 2011; we will consider this just below), is that it cannot account for double conjunction fallacies [1]. So far, there has been limited evidence for double conjunction fallacies (Crupi et al., 2018; Yates & Carlson, 1996; Wojciechowski & Pothos, 2018); we hope to make progress with this important empirical question in the present work. Finally, the probability plus noise model assumes that judgments come from sampling frequencies and thus follow the binomial distribution. In the case of small sample sizes, this means that many possible judgments would be predicted to have a zero likelihood of occurring. To circumvent this limitation, Costello et al. (2014) proposed that people round numbers in a specific way to produce the final probability judgments. Rounding mechanisms have been well supported in the literature and are a reasonable addition to a model for probabilistic reasoning. Nevertheless, it is interesting to consider whether a model with no external rounding mechanism can perform equally well.

The Bayesian Sampler (Zhu, Sanborn, & Chater, 2020) is a related model, based on an analogous sampling process. Like Costello and Watts (2014), the Bayesian sampler assumes that there is a separation between internally generated frequencies and observed participant responses. The Bayesian Sampler model assumes that the sampling process is veridical, subject to sampling limitations, but there is a second step of integrating the sample estimates with a prior belief to produce a final biased mean estimate. In this work we focus on the Bayesian Sampler, because predictions between the two models mostly converge and, where they diverge, the evidence seems to favor the Bayesian Sampler (Zhu et al., 2020). Additionally, a minimal modification of the Bayesian Sampler model, which

[1] In a single conjunction fallacy, the conjunction is rated as more probable than one marginal; with a double conjunction fallacy, the conjunction is rated as more probable than both marginals.

will be discussed in detail later, can solve the likelihood-zero problem of the probability plus noise model mentioned above.

The second main idea in the present work is that the relevant probabilistic calculus may include a non-Bayesian influence. The relevance of quantum theory in cognition can be motivated in a way which is, perhaps surprisingly, very similar to that for the Bayesian Sampler. Zhu et al. (2020; see also Lieder & Griffiths, 2019) note that "We start from the perspective that people, quite possibly implicitly, have an internal Bayesian model of the tasks they engage in. ... A serious challenge to Bayesian models is that Bayesian calculations (e.g., inferring and averaging over the posterior distribution) appear computationally daunting." Note, it may not be immediately obvious why baseline Bayesian theory is intractably complex. One way to see this is in terms of the exponential growth in the complexity of joint probability distributions, as the number of predicates grows. A counterargument might be that if we assume independence then this exponential growth does not occur. However, complete independence is unrealistic and partial independence, e.g., in the form of Bayesian networks, is still problematic (Pothos et al., 2021). In any case, in terms of motivating quantum theory a similar point applies. That is, and as noted, the use of quantum theory in cognition can be motivated as a framework for probabilistic reasoning, which mitigates the computational intractability of baseline Bayesian theory. But how does this come about?

Bayesian and quantum theories are based on different axioms and offer strikingly different approaches to the representation of information and computation of probabilities. Bayesian theory has a set-theoretic structure, so that probabilities are computed against subsets of an overall sample space, while quantum theory is a geometric model of probability, whereby probabilities correspond to projections of a state vector onto different subspaces. For example, in Figure 1a, probabilities for different possibilities about Linda are subsets, whereas in Figure 1b the state vector ψ is projected to different subspaces. As seen in Figures 1a, 1b, the set-theoretic structure of Bayesian theory illustrates the requirement of closure: we are always able to make a judgment about, e.g., overlap (and so probability) for any combination of subsets. By contrast, in quantum theory, incompatible questions are

impossible to resolve concurrently; these are questions for which the corresponding subspaces are at 'angles' to each other, which are not multiples of $\pi/2$. Compatible questions in quantum theory on the other hand are such for which the situation is entirely Bayesian. It is in this way that we have argued that quantum theory may be a plausible bounded rational approach, for agents striving to be Bayesian, but overwhelmed by the multitude of questions they are faced with (Pothos & Busemeyer, 2022): rather than trying to be fully Bayesian for all available questions, an agent aims to do so only for the subsets of compatible questions.

In the context of probability judgments, a quantum probability model can offer an axiomatic way of modeling various probability fallacies such as conjunction and disjunction fallacies, order effects, and violations of law of total probability, in a unified account. Additionally, we are led to a fairly natural way to understand influential distinctions in cognitive theory, such as between slow versus fast, reflective versus reflexive, or analytic versus heuristic thinking (Elqayam & Evans, 2013; Kahneman, 2001). We propose that such distinctions can be understood as ones of Bayesian versus quantum computations. Indeed, there is some evidence that task demands, familiarity, and individual differences can affect the relative weight of Bayesian versus quantum reasoning, as one would expect if the former versus the latter correspond to, broadly speaking, slow versus fast cognition (Trueblood, Yearsley, & Pothos, 2017). In our work, we continue to explore the boundary between quantum and Bayesian cognition, by presenting a model which allows a seamless (parametric) transition between quantum versus Bayesian reasoning.

Busemeyer et al.'s (2011) quantum probability model is incomplete in modeling human probability judgments. In fact, since Busemeyer et al.'s (2011) model is (mostly) just quantum probability theory, it offers little parametric flexibility to accommodate some important findings. For example, Busemeyer et al. cannot explain the violations of some probability identities (Costello & Watts 2014, Costello & Watts 2016). Additionally, Busemeyer et al.'s (2011) model, as well as the Bayesian Sampler model, cannot explain violations of the identity equation, that is $P(A) + P(\neg A) \neq 1$, namely binary complementarity. Note, A here can be either marginals or a more general event (e.g., a conjunction, disjunction, conditional).

Binary complementarity deserves a few further remarks, because of its importance in modeling, as putative violations are beyond the scope of all current theories of probabilistic judgments. Binary complementarity may appear too obvious to be violated and indeed consistency with this constraint is well documented in the literature (Tversky & Koehler 1994; Budescu et al. 1997; Wallsten et al. 1993). Theoretically, binary complementarity is essential in support theory (Tversky & Koehler, 1994). However, prior work already offers some hints that binary complementarity may not always be behaviorally valid. Violations of binary complementarity have been observed in choice behavior (Shafir, 1993; see also Macchi et al. 1999) and in similarity judgements (Tversky & Gati, 1978). Violations of this constraint are related to the subadditivity effect in unpacking, whereby the probability of a packed disjunction event or category is lower than the sum of the probabilities of its unpacked mutually exclusive components (Tversky & Koehler, 1994), especially when the components are typical instances of the packed category (Sloman et al. 2004). More pertinently, Epping and Busemeyer (2023) showed that when one presents $A \wedge \neg A$ as the two only possible alternatives (e.g. Gift Card A vs Gift Card B; there is no other choice), rather than Gift Card A and not Gift Card A, the constraint can be violated. Given the above, whether violations of binary complementarity are present in the present data is an important question we will address.

The third main idea in our work is that there might be a separation between relevant probabilistic principles and the response mechanism. The work of Costello and Watts (2014) and Zhu, Sanborn, and Chater (2020) are recent examples for how to accomplish this. Costello and Watts (2014) assume noisy sampling and Zhu et al. (2020) that the sampling process is veridical (subject to sampling limitations), but there is a second step of introducing bias in responding, when the internal estimates are adjusted against prior beliefs. An important assumption in both these models is that the samples for probabilistic calculations need to be specified in advance. That is, the cognitive agent needs to make a commitment regarding the extent of her sampling, at the initial state of their probabilistic judgment. However, a sampling process for probabilistic reasoning need not be specified in

such a way: notably, sequential sampling models assume that agents collect samples sequentially, with the duration of the sampling process flexibly limited or extended depending on the accumulated evidence, time pressure, engagement with a task, amongst possible factors (e.g., Brown & Heathcote, 2008; Ratliff & Smith, 2015; Trueblood, Brown, & Heathcote, 2014; Usher & McClelland, 2001). In our proposal for probabilistic reasoning, we thus assume that there is a separation between subjective probabilities and response production and employ a sequential sampling process for the latter.

The idea of a separation between internal probabilities and response mechanisms allows us to address an interesting question regarding models of probabilistic reasoning: do probability judgments require or assume rule following by people? As with Costello and Watts (2014) and Zhu et al. (2020), we assume that people's probability judgments are consistent with the rules of formal probability theory, whether Bayesian or quantum. Yet, what people articulate as "probabilities" are typically not pure subjective probabilities but rather 'noisy' judgments influenced by these underlying probabilities. Marr's analysis (Marr, 1982) provides a framework for understanding this distinction: formal probability rules offer a computational-level or top-down (Griffiths et al., 2010) explanation of probabilistic reasoning, whereas a process such as noisy sampling mechanism aligns with Marr's algorithmic-level description, detailing how such inferences are formulated. At the same time, some aspects of probabilistic reasoning may be guided by heuristic rules, rather than formal probabilistic ones, such as the representativeness one from Tversky and Kahneman (1983). Note, as discussed, representativeness is limited in scope; but one can imagine similar principles capturing aspects of probability estimation, outside accounts based on formal probability theory. In any case, if rules are involved in probabilistic reasoning – especially rules from formal probability theory –the relevant computations and cognitive processes are likely to be outside direct conscious control and awareness: lay people, without any mathematical training, are perfectly capable of forming probabilistic intuitions – it is these intuitions we are trying to explain. This is analogous to how young children can perform intuitive physics, without learning classical physics. The consideration of explanation levels as above under-

scores the importance of algorithmic-level models, such as a sampling algorithm that can account for the intuitive generation of responses based on rules.

Overall, the above ideas certainly have much merit. However, it also seems fairly clear that the predominant formalisms for probabilistic reasoning suffer from notable limitations, even in the absence of evaluation against larger datasets, which, we think, are likely to offer additional challenges to existing models. We next describe a novel experiment, to collect an extensive dataset on probabilistic reasoning, and follow with the specific novel theoretical proposals.

Experimental Investigation

Both Costello and Watts (2014) and Zhu et al. (2020) asked participants to judge the probabilities of pairs of weather events. A priori, there are reasonable grounds for expecting that such judgments might be more likely to conform to Bayesian constraints, because we are generally familiar with judgments for weather events, a weather event is less likely to create unique contexts or perspectives for other weather events, and resolving a weather question is unlikely to create an impression of 'disturbance' for subsequent related questions (Pothos & Busemeyer, 2022). Overall, there is little doubt that human judgments are sometimes consistent with Bayesian constraints. Therefore, it is more interesting to examine behavior with judgments more likely to challenge Bayesian prescription. We carried out two pilot experiments to pre-select materials more likely to result in apparent probabilistic fallacies, which are described in Supplementary Material 1. The results of the pilot studies were analyzed primarily in terms of the emergence of conjunction fallacies, without detailed analyses or model fits.

For both the two pilots and the main experiment, we requested judgments concerning the presidential election in the USA in 2020. This presidential election attracted, for various reasons, widespread interest and was extensively covered internationally. Therefore, we anticipated that judgments concerning the probability of the two candidates winning or losing different states would offer an engaging and interesting task to participants, who were

all recruited in the USA.

All experiments were approved by City University London Research Ethics Committee with the ethics approval code ETH2223-0571 and title of study "Decision making for election results". We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. This study was not preregistered; all data and data analysis codes will be avaliable in our github repository: `https://github.com/adamhuang11111/quantum_sequential_sampler_public`.

Main experiment

Participants. We recruited 1451 (908 male) participants, restricting geographical location to the USA, from Amazon Mechanical Turk. Sample size was determined a priori primarily on the basis of practical considerations: recruitment took place just before the presidential election in the USA in 2020 and we recruited the maximum number of participants we believed we could test within a reasonable period prior to the election. No restrictions apart from location were placed on participation. Each participant was paid $2.25 and the experiment lasted approximately 25 minutes. A simple attention check question was included: halfway the survey, a question similar in style to the other ones was presented, asking participants to simply move the slider to a specific number (e.g., 47). As a result of failing to answer the attention check question correctly, being identified as a spam bot, or having provided incomplete data, 289 participants were excluded from further analysis. Thus, the final sample size was reduced to 1162 participants (730 male). 1118 out of these participants were at least 25 years of age and therefore eligible to vote in the USA.

Method. Participants were asked to provide 78 probability judgments (below, we invariably refer to these as probability questions, events, or judgments) concerning the likelihood of one or both presidential candidates (Trump and Biden) winning the popular vote in the states corresponding to the triplets chosen from the pilot experiments:

- $T1$: Ohio, Missouri, Michigan
- $T2$: Georgia, Montana, Nevada

1	A	marginal	15	$\neg A \vee B$	disjunction order 1
2	B	marginal	16	$\neg A \vee \neg B$	disjunction order 1
3	$\neg A$	marginal	17	$B \vee A$	disjunction order 2
4	$\neg B$	marginal	18	$\neg B \vee A$	disjunction order 2
5	$A \wedge B$	conjunction order 1	19	$B \vee \neg A$	disjunction order 2
6	$A \wedge \neg B$	conjunction order 1	20	$\neg B \vee A$	disjunction order 2
7	$\neg A \wedge B$	conjunction order 1	21	$A\|B$	conditional
8	$\neg A \wedge \neg B$	conjunction order 1	22	$A\|\neg B$	conditional
9	$B \wedge A$	conjunction order 2	23	$\neg A\|B$	conditional
10	$\neg B \wedge A$	conjunction order 2	24	$\neg A\|\neg B$	conditional
11	$B \wedge \neg A$	conjunction order 2	25	$B\|A$	conditional
12	$\neg B \wedge \neg A$	conjunction order 2	26	$B\|\neg A$	conditional
13	$A \vee B$	disjunction order 1	27	$\neg B\|A$	conditional
14	$\neg A \vee B$	disjunction order 1	28	$\neg B\|\neg A$	conditional

Table 1: The 28 probability judgments required for the pair of events {A, B}.

If we label the states in a triplet as A, B, C, Table 1 shows the 28 probability judgments for pair A, B. For pair A, C, there are an additional 26 judgments because the marginals for A are already covered in the first set of 28 judgments; and for pair B, C, an additional 24 judgments, for a total of 78. Note, we can arbitrarily consider P(Biden to win A) as $P(A)$ and P(Trump to win A) as $P(\neg A)$, thereby avoiding the need to ask participants to rate event negations. Even though there were other candidates, Biden and Trump were the dominant ones and, as an approximation, we can ignore the possibility of other candidates; no other candidates were mentioned in the experiment. The judgments were comprised of six marginals, 12 conjunctions, 12 disjunctions and 12 conditionals, with all composite events presented in both possible orders. All participants were asked to rate the marginal probabilities first, before being presented questions about composite events in different blocks (described just below). The reason why the marginals were shown first was so that participants would be exposed to the range of atomic events, prior to any other questions. Within blocks, question order was randomized.

The rating scale consisted of an adjustable slider, with anchor points 0% and 100% and movement in 1% increments. The slider consisted of a circle, which participants could move around with a mouse. Just above this circle, participants could see the rating the

slider corresponded to, at any given position. Additionally, above the slider, we indicated the locations of 10% increments. The slider was always initialized at 50 in all trials. Note, this might be a source of bias, for example, in that more effort would be required to produce more extreme responses. There are two mitigating considerations. First, simple inspection of the probability judgments distributions (Figure 7) shows that the mode of many of these distributions is away from 50. This indicates that, even if there is a bias, it is not strong enough to dominate the mode of the distributions. Second, as will be explained later, there is an overestimation bias present in the probability judgments. Should a motor bias be influencing these judgments, due to the initial placement of the slider at its midpoint, we would expect to see a more balanced distribution of judgments across the slider's range. Supplementary Material 1 offers more details about the procedure for rating elicitation and some example screenshots of the slider we employed.

The design involved two between participants conditions. The first condition was the triplet, triplet 1 ($T1$) versus triplet 2 ($T2$). We tested different participants on each triplet, as a way to limit the total number of probability judgments for each participant. The second condition was a counterbalancing one, corresponding to whether participants completed all judgments for a particular pair first before proceeding to the judgments for another pair (blocked order, BO) versus completing all judgments for all pairs together, in a randomized order (fully randomized order, FO). The number of participants in each combination of conditions was, for the $T1$ BO, $T1$ FO, $T2$ BO, $T2$ FO conditions, 284, 301, 269, 308, respectively. As participants in the BO condition did not provide responses noticeably different from those in the FO one, this counterbalancing condition will not be further considered.

In both the BO and the FO conditions, probability judgments were blocked by type of judgment, so that when participants completed the judgments for one block there was a small break, before proceeding to the next one. In the FO case, participants first responded to all possible conditionals in one direction (e.g., $A|B$), then conjunctions, then disjunctions. Subsequently, participants saw the same judgments in the reverse direction

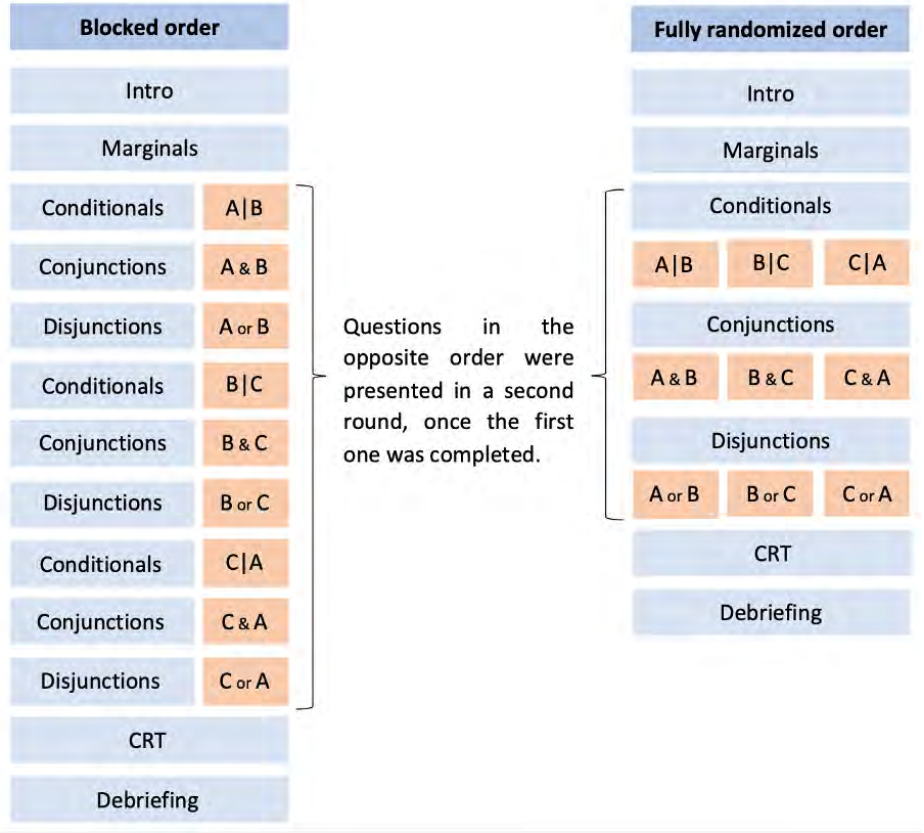


Figure 2. Survey flow highlighting the differences between the Blocked order (BO) and the Fully randomized order (FO) conditions.

(e.g., $B|A$). The BO case was analogous, but the judgments corresponding to each pair of states were also blocked. Within each block, judgment order was randomized for the first presentation, but kept the same for the second presentation. Once participants completed all probability judgments, they were asked to answer three questions corresponding to the Cognitive Reflection Test (CRT, Frederick, 2005). They were then debriefed, thanked, and paid for their participation. Figure 2 provides a sketch of the main parts of the survey flow in the experiment.

Behavioral analyses

Response biases. We first considered participants' use of the ratings scale, notably whether participants made full use of the ratings scale and whether particular ratings might

have been preferentially employed, e.g., as a result of rounding behavior (Budescu, Weinberg, and Wallsten, 1988; Wallsten, Budescu, and Zwick, 1993). Given there were 78 probability estimates in the experiment, each participant can give a maximum of 78 different ratings. The number of points on the ratings scale which were used by participants varied between 2 and 56 (mean = 31.45, std = 11.65). Only 29 participants out of 1162 used fewer than 10 different points on the rating scale, demonstrating that most participants made reasonable use of the rating instrument provided. In Supplementary Material 2, Figure S.2.1 shows how often each rating was observed in participant responses, with the bars corresponding to multiples of 5 highlighted. It appears that such ratings were indeed preferentially employed, but to a lesser extent, compared to that in Zhu et al. (2020), where probability judgments were numbers which were entered into a computer.

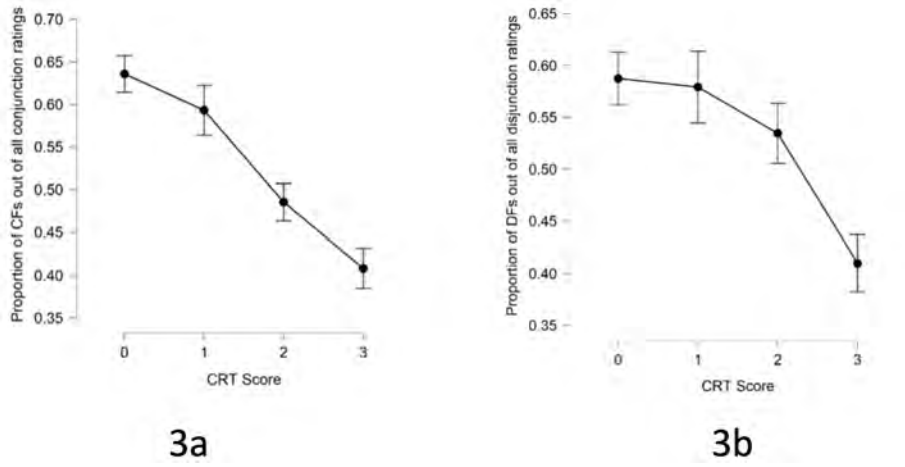


Figure 3. An illustration of the relationship between CRT and conjunction (3*a*) and disjunction (3*b*) fallacies.

Conjunction fallacy. Single versus double conjunction fallacies can be identified by considering whether the conjunction is higher than one or both marginals, respectively. There was a high rate of conjunction fallacies in the dataset, with 59.4% of all conjunctions associated with a conjunction fallacy. Of these cases, 38% corresponded to single conjunction fallacies and 62% to double conjunction fallacies; assessed against the total number of conjunctions these percentages were 22.7% and 36.7% respectively. Therefore, our work pro-

Identity name	**Identity calculation**
Z_1	$P(A) + P(B) - P(A \cap B) - P(A \cup B)$
Z_2	$P(A) + P(B \cap \neg A) - P(B) - P(A \cap \neg B)$
Z_3	$P(A) + P(B \cap \neg A) - P(A \cup B)$
Z_4	$P(B) + P(A \cap \neg B) - P(A \cup B)$
Z_5	$P(A \cap \neg B) + P(B \cap A) - P(A)$
Z_6	$P(B \cap \neg A) + P(A \cap B) - P(B)$
Z_7	$P(A \cap \neg B) + P(B \cap \neg A) + P(A \cap B) - P(A \cup B)$
Z_8	$P(A \cap \neg B) + P(B \cap \neg A) + 2P(A \cap B) - P(A) - P(B)$
Z_9	$P(A\|B)P(B) - P(B\|A)P(A)$
Z_{10}	$P(A\|B)P(B) + P(A\|\neg B)P(\neg B) - P(A)$
Z_{11}	$P(B\|A)P(A) + P(B\|\neg A)P(\neg A) - P(B)$
Z_{12}	$P(B\|A)P(A) + P(A\|\neg B)P(\neg B) - P(A)$
Z_{13}	$P(A\|B)P(B) + P(B\|\neg A)P(\neg A) - P(B)$
Z_{14}	$P(A\|\neg B)P(\neg B) + P(B) - P(B\|\neg A)P(\neg A) - P(A)$
Z_{15}	$P(A \cap B) - P(A\|B)P(B)$
Z_{16}	$P(A \cap B) - P(B\|A)P(A)$
Z_{17}	$P(A \cap B) - P(A) + P(A\|\neg B)P(\neg B)$
Z_{18}	$P(A \cap B) - P(B) + P(B\|\neg A)P(\neg A)$

Table 2: Probabilistic identities from (baseline) Bayesian theory, according to Zhu, Chater and Sanborn (2020). In all cases, the predicted value is 0. Note: Identities are abbreviated using $P(\neg A)$ and $P(\neg B)$ for $1 - P(A)$ and $1 - P(B)$.

vides new evidence that double conjunction fallacies can arise in human judgments (Crupi et al., 2018; Yates & Carlson, 1986 ; Wojciechowski & Pothos, 2018).

We examined whether the emergence of probabilistic fallacies, such as the conjunction fallacy, can be tied to individual differences concerning the CRT, which has been employed in similar ways in the past (Yearsley & Trueblood, 2018). We approximated with CF_{rate} (Supplementary Material 1) the number of instances in which a conjunction fallacy, regardless of size, was detected per participant. Note, results here were analyzed for both triplets together; for this analysis, separating out results from each triplet is irrelevant. A one-way between participants ANOVA was conducted, with the proportion of conjunction fallacies as the dependent variable and the CRT score as an independent variable with four levels, which were the four possible scores in the CRT. A significant effect of CRT scores on the proportion of conjunction fallacies was found, $F(3, 1158) = 78.76, p < .001 (BF_{10} > 150)$. Repeated contrasts revealed that for each point gained in the CRT score, the frequency

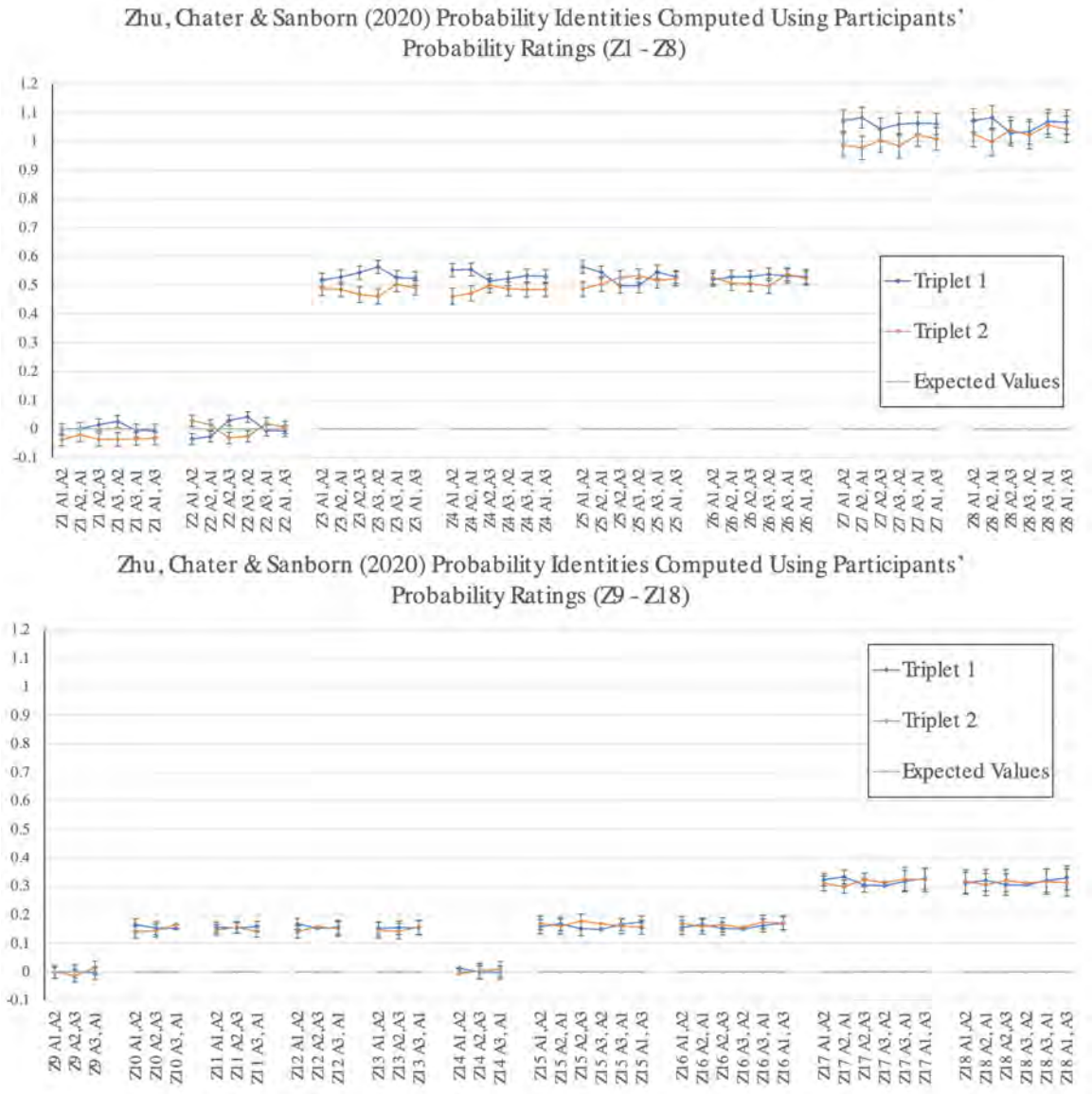


Figure 4. The observed values for the Z-identities (Table 2) in the present data set, computed from average probabilities, separately for each of the two triplets and orders of conjunctions and disjunctions. In all cases, the predicted value is 0.

of conjunction fallacies was reduced ($p = .001 - .017$, BF_{10} ranging from 1.27 to > 150). A summary of the contrast analyses can be found in Table S2.1 and an illustration of the relationship between conjunction fallacies and CRT is in Figure 3.

Disjunction fallacy. Analogous to conjunction fallacies, a disjunction fallacy occurs when the probability of a disjunction is judged lower than that of either constituent. In the present dataset, there was an apparent disjunction fallacy for 53% of all disjunctions. Most disjunction fallacies corresponded to a single disjunction fallacy (63.9%), but there was a sizeable proportion of double ones as well (26.1%).

The proportion of disjunction fallacies DF_{rate} (Supplementary Material 1), that is the number of instances in which a disjunction fallacy (again regardless of size) was detected, differed across participants with different CRT scores, as assessed with a one-way between participants ANOVA, with CRT as a four-level independent variable, $F(3, 1158) = 31.709, p < .001 (BF_{10} > 1000)$. As seen in Figure 3, the rate of disjunction fallacies is reduced with increasing CRT, a pattern that was mostly confirmed statistically, with repeated contrasts: the disjunction fallacy rate of participants with a CRT score of 3 was lower than that of all other participants, $p < .001 (BF_{10} > 106)$; analogously for participants with a CRT of 2 relative to ones with a CRT score of 0 ($p = .029 (BF_{10} = 3.11)$, but no other comparisons were significant (Table S2.2). [2]

Order effects. Any deviation between the conjunction in one order and the same conjunction in the opposite order, regardless of how small it is, implies the presence of an order effect, that is, a conjunction order effect is evidenced when $P(A \text{ and then } B) \neq P(B \text{ and then } A)$ and a disjunction order effect when $P(A \text{ or then } B) \neq P(B \text{ or then } A)$. It is notable that such order effects are beyond both Bayesian theory and the current quantum models for the conjunction fallacy. In the latter case, in quantum theory, it is possible to have $P(A \text{ and then } B) \neq P(B \text{ and then } A)$. However, in the original quantum model, it was assumed that conjunctions are evaluated so that the more likely predicate is considered first, regardless of the order in which the predicates appear in the conjunction (analogously for disjunctions; Busemeyer et al., 2011). Instead, order effects have been postulated to be relevant when participants answer one question after another (Wang et al., 2014). So,

[2] In Table 2 and throughout the paper, we use $\cap$ and $\cup$ to denote classical probabilities, and $\wedge$ and $\vee$ to denote probabilities with "and" and "or" that do not need to obey clasical probability theory like those found in the probability judgments and quantum probabilities.

if this approach is behaviorally accurate, then we do not expect systematic order effects in conjunctions and disjunctions. The results, shown in Tables S2.3 and S2.4, indicate no evidence for order effects, based on Bayesian one sample t-tests against 0, in nearly all cases.

Concerning a putative association between order effects and CRT, the absolute magnitude between the same conjunction or disjunction, in different orders, can be taken as a measure of deviation from both classical prediction and the prediction from Busemeyer et al.'s (2011) quantum model. A between-subjects ANOVA with the average order effect size as the dependent variable and the CRT score as the independent variable (with four levels) revealed an effect of CRT score on order effect size, $F(3, 1158) = 3.045, p = .028$. However, post-hoc t-tests failed to show systematic changes in effect size depending on the CRT score.

Reciprocity. The constraint of reciprocity is that $P(X|Y) = P(Y|X)$. Clearly, reciprocity does not apply to Bayesian theory and it does not apply to quantum theory, when subspaces of varying dimensionalities are employed (Busemeyer et al., 2011). However, quantum models with one-dimensional subspaces are constrained by reciprocity (Busemeyer & Bruza, 2011) and such models are sometimes employed (e.g., White, Pothos, & Jarrett, 2020; Yearsley & Pothos, 2016). There is some evidence that humans are sometimes constrained by reciprocity (Trueblood et al., 2017). In the present data, overall, there was limited evidence for reciprocity. Out of the 78 probability judgments, there were 24 conditional probability ones (12 pairs for each triplet, so 24 pairs in total). Only in 6/24 pairs of matched conditional probability judgments (e.g., $P(A|B)$ versus $P(B|A)$) was there evidence for reciprocity, using Bayesian paired-samples t-tests (Table S2.5).

Z-identities. Costello and Watts (2016) derived several probabilistic identities that must hold in a baseline Bayesian probability framework. The list of these identities was also tested by Zhu, Sanborn and Chater (2020; Table 2). If participant judgments can be described by basic Bayesian probabilities, that is without noise or other additional assumptions, one would expect all identities Z_1 to Z_{18} to be equal to zero. Violations of these identities were found in the experiments of Costello and Watts (2016) and Zhu, Sanborn and Chater (2020), indicating, unsurprisingly, that peoples' judgments are not consistent

with baseline Bayesian theory.

For the present data, the Z-identities were tested using average probabilities from participant ratings within each triplet. Each test was based on a Bayesian one-sample t-test, against zero. Systematic deviations from zero could be observed for all identities, apart from Z_1, Z_2, Z_9 and Z_{14}. The observed Bayes factors indicate that the alternative hypothesis, $Z \neq 0$, is vastly more likely than the null (Table S2.6; Figure 4). For the four identities for which evidence was consistent with the Bayesian expectation, $Z_{1,2,9,14} = 0$, evidence ranged from anecdotal to strong (BF_{01} from 1.2 to 9.1). This pattern of results differs in an interesting way compared to the one in Zhu et al. (2020). Notably, we observed higher values for many of these identities, compared to Zhu et al. (2020), which raises the question of whether the Bayesian Sampler model will cope with the present results. We compared the identity values from the present dataset vs the ones from Zhu et al.'s (2020) dataset (Supplementary Material 7). There was strong evidence for differences in all cases, except from Z_1, Z_9, Z_{15}, Z_{16}.

Binary complementarity and the law of total probability. A key requirement for both Bayesian and quantum theories concerns binary complementary and the law of total probability. Note, for quantum theory, interference effects can allow violations of this constraint only when conjunction orders are mixed. Specifically, we expect consistency with binary complementarity, such as $P(A) + P(\neg A) = 1$ and the four-way law of total probability $P(A \cap B) + P(A \cap \neg B) + P(\neg A \cap \neg B) + P(\neg A \cap \neg B) = 1$. The conjunction fallacy and related findings perhaps encourage an expectation for violations regarding expressions including conjunctions or disjunctions. However, in any specific dataset, it cannot be taken for granted than such violations will emerge and, moreover, it is unclear what to expect for the seemingly obvious version of binary complementarity with marginals.

We tested binary complementarity for marginals and the four-way law of total probability in the present dataset using one-sample Bayesian t-tests (against 1). These tests yielded strong evidence for deviations from 1 for both equations, that is none of the identities hold for the present data: $BF_{10} > 10^{50}$ for all identities, regardless of the particular

triplet (triplet 1 or triplet 2) or the order of conjunctions (cf. Erev, Wallsten, & Budescu, 1994). The results of the Bayesian t-tests are shown in Table S2.7 and the distributional information for the various versions of the identities, depending on which event or pair of events is considered, in Figure S.2.2.

The observed violation of binary complementarity, $P(A) + P(\neg A) > 1$, might be related to the unpacking effect (Tversky &Koehler, 1994). In the present case, even though Trump and Biden were the two only plausible alternatives for presidential candidates, presenting these typical, unpacked events of the packed category "presidential candidate" may be inducing a subadditivity unpacking effect, leading to overestimation. In any case, violations of binary complementarity are a surprising, important finding. Note, the effect is unidirectional: in our dataset 72.1% of all binary complementarity expressions produce more than 1, i.e., there is a reliable overestimation effect. As far as we know, because this is a relatively newly discovered effect in probabilistic reasoning, no model can directly account for this overestimation effect for binary complementarity, other than the Quantum Sequential Sampler. Behaviorally, the most surprising violation of binary complementarity concerns marginals, but there are several versions all trending in the same direction (e.g. $P(A \cap B) + P(\neg A \cup \neg B) > 1$). Figure 9 provides more details about violations of binary complementarity for all possible complementary pairs.

Given how surprising violations of binary complementarity are, especially for marginals, we can ask whether this effect might be due to some aspect of the data collection procedure. We followed all the standard procedures and incorporated several checks to ensure data quality. Note, the procedure we adopted closely follows Zhu et al. (2020). There are three further considerations mitigating any concerns about the validity of the observed violations of binary complementarity. First, all the marginals were presented together, prior to the more complex probabilities. As noted, we chose this method to make participants aware of all the basic events early on. Moreover, this meant that participants would see events and their negations in close proximity, suggesting that any violation of binary complementarity do not arise from, for example, memory failures. Second, regarding

binary complementarity for marginals, events and their negations were not presented explicitly as such, e.g., as $P(A)$ versus $P(\neg A)$. Instead, participants would see e.g., P(Biden to win in state X) versus P(Trump to win in state X). Plausibly, this encouraged violations of binary complementarity (as in Epping & Busemeyer, 2023). Third, the empirical probability judgments were collected using a slider that defaults to a neutral midpoint of 50 (see Supplementary Material 1). Considering this neutral starting point, it seems implausible to attribute the overestimation effect to the data collection instrument itself.

Generally, there is a long history of surprising and counterintuitive findings in probabilistic reasoning. The observed violation of binary complementarity in the present study is indeed puzzling, given the intuitiveness of this constraint and the broad spectrum of literature supporting it (Tversky & Koehler 1994, Budescu et al. 1997, Wallsten et al. 1993). However, the intuitiveness of a probability constraint should not get in the way of rejecting it, when there is compelling and statistically significant empirical evidence. Indeed, we could easily imagine a reviewer of Tversky and Kahneman (1983) asking 'how can we take seriously data where participants judge a conjunction as more probable than a marginal, when the two judgments are so close to each other?'. Invariably, the key drivers of decision research have exactly been surprising and unexpected findings like these.

Computational Models

Bayesian Sampler

Zhu et al. (2020) proposed that probability judgments are generated from Bayesian reasoning, based on subjective probabilities estimated from an internal sampling process and a biased prior distribution. Specifically, the Bayesian Sampler model assumes that participants initially have a symmetric beta prior distribution of probability judgments, which is updated, using Bayes rule, via mental sampling. A response to a probability judgment would then correspond to the binomially distributed expected values of the beta posterior distribution after mental sampling. The sample size of the mental sampling process is considered a free parameter of the model. Since responses depend on both the frequencies

for mental sampling and the prior distribution, the Bayesian Sampler has the flexibility to generate various probability fallacies. Note, concerning conjunction and disjunction fallacies, Zhu et al. (2020) required two further assumptions, which are analogous to the ones made by Costello and Watts (2014): first, that such judgments are computationally more expensive to simulate and thus employ smaller mental sample size; second, that different samples are employed for e.g. conjunctions and corresponding marginals.

Zhu et al.'s (2020) model shares some similarities with Costello and Watts' (2014) model and there is overlap in predictions too, for example, the Bayesian Sampler model also adjusts probabilities away from extreme values. Zhu et al. (2020) reported an equation linking the noise parameter in Costello and Watts' (2014) model with the sample size parameter in the Bayesian Sampler. Formal comparison between the probability plus noise model and the Bayesian Sampler is complicated by the fact that the two models make equivalent average predictions for probabilities of single events, conjunctions, and disjunctions. However, the two models diverge for conditional probabilities and the distribution of probability estimates and, on that basis, Zhu et al. (2020) concluded in favor of the Bayesian Sampler model. So, in the present work we focus on the Bayesian Sampler. In the following, we summarize the mathematical details of the Bayesian Sampler model and discuss how we fitted the model to the present data.

The model assumes that previous experience establishes a generic prior on probability judgments, with a symmetric beta distribution, $Beta(\beta, \beta)$. As noted, this prior is updated in light of information from an internal sampling process, according to Bayes rule. Let N be the sample size in this internal process. Then, $S(A) \sim Bin(N, P(A))$ and $F(A) = N - S(A)$ denote the instances consistent and not consistent with some generic event A occurring in the sampling process, assuming that $P(A)$ is the subjective probability of A (A can be a marginal, a conditional, a conjunction etc.). The posterior distribution for A, given a particular sample in which there are $S(A)$ instances consistent with A, has the form:

$$P_{BS}(A|S(A)) \sim Beta(\beta + S(A), \beta + F(A)). \quad (1)$$

This approach can be readily adapted to identify the posterior probability correspond-

ing to conditional event A|B, so that $P_{BS}((A|B)|S(A|B)) \sim Beta(\beta+S(A|B), \beta+F(A|B))$, where now $S(A|B) \sim Bin(N, P(A|B))$ denotes the number of times event A occurs, in a sample of size N where B is true; as before, $F(A|B) = N - S(A|B)$. To account for the conjunction and disjunction fallacies, the Bayesian Sampler assumes that $N' \leq N$, where N' denotes the number of samples to evaluate a conjunction or disjunction and N the samples for any other judgment.

The posterior distributions for the conjunctions and the disjunctions follow from Equation (1) above, with the various quantities defined analogously, e.g., $S(A \wedge B)$ is the number of instances in the sampling process whereby both A and B occur and $S(A \wedge B) \sim Bin(N', P(A \cap B))$.

$$P_{BS}(A \wedge B|S(A \wedge B)) \sim Beta(\beta + S(A \wedge B), \beta + F(A \wedge B)), \tag{2}$$

$$P_{BS}(A \vee B|S(A \vee B)) \sim Beta(\beta + S(A \vee B), \beta + F(A \vee B)). \tag{3}$$

Zhu et al. (2020) assume that the reported estimate for the probability of individual event A is the mean of the corresponding posterior distribution (Equation 1). The mean of the posterior distribution of probability judgments (Equation 1), for a specific value of $S(A)$, is given by

$$E[P_{BS}(A|S(A))] = \frac{S(A) + \beta}{N + 2\beta}. \tag{4}$$

Note that $P_{BS}(A)$ is binomially distributed as $S(A)$ follows the binomial distribution. The assumed reported probabilities for conditional events, conjunctions, and disjunctions follow from Equation (4) and are given by (note, $N' \leq N$, for conjunctions and disjunctions):

$$\begin{aligned} E[P_{BS}((A|B)|S(A))] &= \frac{S(A|B) + \beta}{N + 2\beta}, \\ E[P_{BS}(A \cap B|S(A \cap B))] &= \frac{S(A \cap B) + \beta}{N' + 2\beta}, \\ E[P_{BS}(A \cup B|S(A \cup B))] &= \frac{S(A \cup B) + \beta}{N' + 2\beta}. \end{aligned} \tag{5}$$

For a particular probability judgment, if the sample size is N, the Bayesian Sampler makes N discrete predictions, distributed binomially, as the means of the beta posterior

distribution (recall, the mean of the beta posterior is itself a random variable). If for a probability judgment N is small, say 5, the Bayesian Sampler predicts only five possible responses, e.g., 0.11, 0.20, 0.30, 0.40, 0.50. Then, if a participant's responses for this judgment is e.g. 0.12 (only 0.01 away from one of the predicted responses), the likelihood of this response from the model is 0. This is a counterintuitive and unrealistic constraint.

Zhu et al. (2020) circumvent this zero-likelihood issue with two methods. First, they fitted only the means of the repeated measurements from the same participant and compared it with the expected value of posterior means, using a sum of squares error. However, this approach under-represents data distributional information. Second, they proposed an external rounding mechanism. In their experiment, participants were asked to type answers into the computer corresponding to probability estimates. It is plausible that such reports were influenced by people's tendency to round to the nearest multiple of 0.05, when probabilities are measured in a 0 to 1 scale (Budescu, Weinberg, & Wallsten, 1988; Wallsten, Budescu, & Zwick, 1993). It may be reasonable or not for the Bayesian Sampler to include this additional mechanism, but either way it is interesting to consider whether good fits are possible without it. Note, in our study, we asked participants to report probability values using a continuous rating slide, so that a rounding mechanism to the nearest multiple of 0.05 would not have been evoked to the same extent as in Zhu et al. (2020).

In order to fit the Bayesian Sampler model by maximum likelihood to the continuous data we obtained, we circumvented this zero-likelihood problem for the Bayesian Sampler by a different method: we minimally extended the Bayesian Sampler model, by taking a step back and directly using the posterior beta distribution corresponding to the predictions for a probability judgment. We can then assume that people report a sampled value from this posterior distribution, instead of reporting the posterior mean. Since the beta posterior is a continuous distribution, the Bayesian Sampler can predict a non-zero likelihood for any empirical judgment. Formally, let $B_{((\beta,S(E),N))}(x)$ denote the probability density function of the posterior distribution $Beta(\beta + S(E), \beta + F(E))$, where E is any possibility for an individual event (a conditional, a conjunction, or a disjunction), and $S(E), F(E)$ are defined

as the samples consistent and inconsistent with E. We also have $S(E) \sim Bin(N, P(E))$ and denote $Bin_{((N,P(E)))}(x)$ as the probability of obtaining $S(E)$ true instances when sampling E events, from a binomial distribution with sample size N. The likelihood of observing a probability judgment of value x from the Bayesian Sampler model (with parameters $N, P(E)$), is given by:

$$\begin{aligned} L_{BS}(x|N, P(E)) &= \sum_{S(E)=0}^{N} P(S(E)) \cdot P(x|S(E)) \\ &= \sum_{S(E)=0}^{N} Bin_{(N,P(E))}(S(E)) \cdot B_{(\beta,S(E),N)}(x). \end{aligned} \tag{6}$$

This likelihood function works for continuous probability judgments, but in our empirical investigation, probability judgments were measured as integers from 0 to 100. We therefore need to discretize the symmetric beta distribution. Specifically, we define a function u that maps integers from 0 to 100 to the beta probability density as:

$$\begin{aligned} u_{(\beta,S(E),N)}(i) = B_{(\beta,S(E),N)}(\frac{i}{100}), 1 \leq i \leq 99, \\ u_{(\beta,S(E),N)}(0) = B_{(\beta,S(E),N)}(0.005), \\ u_{(\beta,S(E),N)}(100) = B_{(\beta,S(E),N)}(0.995). \end{aligned} \tag{7}$$

In Equation (7), the probability density function of the symmetric beta distribution is undefined at exactly 0 and exactly 1, so we approximate the corresponding densities with 0.005 and 0.995. The likelihood function of the Bayesian Sampler model that will be fitted to our data for each event E can then be written as:

$$\begin{aligned} L_{BS}(x|N, P(E)) &= \sum_{S(E)=0}^{N} P(S(E)) \cdot P(x|S(E)) \\ &= \sum_{S(E)=0}^{N} Bin_{(N,P(E))}(S(E)) \cdot \frac{u_{(\beta,S(E),N)}(x)}{\sum_{i=0}^{100} u_{(\beta,S(E),N)}(i)}. \end{aligned} \tag{8}$$

Note, when we shortly present our own Quantum Sequential Sampler model, we will also employ a similar assumption of a symmetric beta distribution for the initial state and the same discretization techniques as in Equation (7). Regarding the Bayesian Sampler,

equation (8) allows us to fit the model by maximizing the product of likelihoods for all of the probability judgments, for each participant.[3] The likelihood value can then be converted to Bayesian Information Criterion (BIC) values.

There are some additional, fairly minor, considerations, before the Bayesian Sampler model can be applied to the present dataset. Recall, the present dataset involves all possible probability judgments, arising from the three pairs formed by the three events we employed. For example, for the three events Biden wins New Hampshire, Trump wins Florida, and Biden wins Penn, we would consider all probabilities from the three pairs Biden wins Penn, Biden wins New Hampshire; Biden wins Penn, Trump wins Florida; and Biden wins New Hampshire, Trump wins Florida. Instead, Zhu et al. (2020) considered probability queries from a single pair of weather events, e.g., normal weather, typical weather. With a single pair of events, there are at most two sample size parameters, corresponding to the sample size employed for estimating marginals and conditionals versus conjunctions and disjunctions. With three pairs of events, we decided to test two variants of the Bayesian Sampler model. With the first variant, we assumed that marginals/ conditionals are sampled using one sample size, N_1, and three further samples sizes were required for the three pairs of conjunctions/ disjunctions, $N_2, N_3, N_4 < N_1$. With the second variant, there was a sample size for marginals/ conditionals, N_1 and a single sample size for conjunctions/ disjunctions, N_2. The two versions of the model are nested and so can be compared through a G^2 test over all participants. The result showed that only 6 out of 1162 participants are fitted significantly better (p value $< .05$) by the more elaborate version of the model, representing $0.5\% < 5\%$ of all participants. Therefore, we fail to reject the simpler model version over all participants and, in further analyses, it is the simpler version of the Bayesian Sampler model which we will consider.

To summarize, the simpler version of the Bayesian Sampler model that we employed has a total of nine parameters, which are {N_1, N_2, $P(A)$, $P(B)$, $P(C)$, $P(B|A)$, $P(C|A)$,

[3]The fitting result following a maximum likelihood approach might be different from that based on minimizing the sum of square error, as performed in Zhu et al. (2020), because the maximum likelihood technique minimizes a weighted sum of square error, that equals the sum of square error (from corresponding means) only when the distribution is normal.

$P(C|B)$, $\beta\}$, where the six subjective probabilities are employed to compute all other probabilities, β is the beta distribution parameter, and the two sample sizes correspond to the marginals/ conditionals and conjunctions/ disjunctions respectively.

Zhu et al. (2023) recently extended the Bayesian Sampler model, by relaxing the assumption that the samples involved in the generation of probabilities are independent; instead, they assumed autocorrelated samples, in their Autocorrelated Bayesian Sampler. Their model was argued to be consistent with a range of findings in probabilistic reasoning, including response times and confidence intervals. However, Zhu et al. (2023, p.12) do note that the Autocorrelated Bayesian Sampler produces average probability judgments approximately equivalent to that of the Bayesian Sampler, except for effects explained by autocorrelated sampling, such as an implicit unpacking effect, which is not assessed in our dataset. Therefore, in this work, it should suffice to compare the Bayesian Sampler with the Quantum Sequential Sampler. Additionally, because the Quantum Sequential Sampler includes a sequential sampling component, it should be possible to extend its application to response times, confidence intervals etc. and so compare with the Autocorrelated Bayesian Sampler – but this is an objective for future work, not least because the Autocorrelated Bayesian Sampler is not yet at a form that can be directly fitted to data.

Quantum Sequential Sampler

As for the Bayesian Sampler model, our proposal of the Quantum Sequential Sampler assumes that probability responses are the result of an internal sampling process. However, the corresponding subjective probabilities are quantum and the sampling process a sequential sampling one. Specifically, the Quantum Sequential Sampler combines Busemeyer et al.'s (2011) axiomatized explanation of probability fallacies with a psychologically plausible response process. The response process is inspired by sampling models but takes a step forward: instead of assuming sampling with fixed sample sizes, we consider a dynamical sequential sampling process. In short, our model assumes that discrepancy from Bayesian reasoning can arise in two ways. First, in a way analogous to that of Costello and Watts (2014) or Zhu et al. (2020), it can arise from the response process. Second, the subjective

probabilities themselves can be more Bayesian or less Bayesian, to varying degrees. That is, we assume that there is a duality of human reasoning, between something which approximates Bayesian reasoning and another influence - our argument is that this alternative influence can be captured by quantum theory. Each individual is not necessarily Bayesian or quantum in a black and white manner, rather there is a continuum covering all intermediate points, from strongly Bayesian to strongly quantum.

There are several theoretical motivations for the present approach. First, Busemeyer et al.'s (2011) quantum model, based nearly exclusively on just the probabilistic calculus from quantum theory, is overly restrictive. Second, Zhu et al.'s (2020) assumptions that sampling complexity varies between conjunctions/ disjunctions versus other probabilities and that samples are drawn independently for each probability judgment are not ideal. Third, it would be desirable to avoid reliance on a rounding mechanism (as in the Bayesian Sampler) and also develop a new model with a dynamical component, with potential for additional predictions such as concerning response times. Finally, the Bayesian sampler and other sampling models have the zero-likelihood problem as mentioned previously. The Quantum Sequential Sampling model, by combining quantum probability with a sequential sampling response process, circumvents these problems.

In what follows, we introduce the Quantum Sequential Sampler model in detail. Our explanation of the model is divided into two parts. In the first part, we explain the quantum internal probabilistic calculus for computing subjective probabilities. In the second part, we introduce the Markov sequential sampling model employed to map subjective probabilities into responses.

Quantum Sequential Sampler first part – subjective probability . Before presenting formal details, it may be helpful to offer a brief example of how quantum probability works, with reference to Tversky and Kahneman's (1983) conjunction fallacy example. In fact, it is helpful to first consider how Bayesian probability theory works. In Figure 1a, we present a classical sample space, which shows a sample of hypothetical Lindas that we can imagine or have experienced (i.e., women like Linda). The red dots are Lindas consistent with the

feminist property, which are numerous, since Linda was described to look like a feminist. Analogously, the blue dots represent instances for which the bank teller property is true. The instances for which both the feminist and the bank teller properties are true are then the intersection of the feminist and bank teller one, shown as dots which are both blue and red, in the smudged area. Clearly, the instances in the intersection can never be more numerous than the instances in either the bank teller or feminist sets and so, in Bayesian theory, it is impossible to have $P(F \wedge BT) > P(BT)$. This example alludes to the set-theoretic or Kolmogorov instantiation of classical probability theory, but there are alternative approaches to formulate classical probability theory (e.g., Cox, 1961, Jaynes, 2003). All the different approaches share deep equivalences, at least insofar that they are all constrained in similar ways.

Figure 1*b* shows a quantum theory caricature of the Linda situation. The state vector ψ represents the mental state after reading the Linda story. Different subspaces correspond to different questions, such that subspace dimensionality reflects the complexity (or facets) of the corresponding question (Pothos, Busemeyer, & Trueblood, 2013). In Figure 1*b*, we show the subspaces corresponding to the answers to the Linda questions as one-dimensional. The state vector is placed close to the feminism subspace and further away from the bank teller one. This is because probability depends on the overlap between the state vector and the corresponding subspace. When a question is resolved, with a probability depending on overlap, the state vector is 'projected' in one of possible subspaces (using a projector). The feminist and bank teller questions are called incompatible, because in most cases we cannot concurrently resolve them. Incompatibility is unique to quantum theory. For incompatible questions, certainty about one question in general implies uncertainty about the other. Quantum theory also allows compatible questions for which the situation is Bayesian. For incompatible questions, conjunctions have to be computed in a sequential way – there is no other possible way to compute such conjunctions (Busemeyer et al., 2011, 2015; Pothos et al., 2017). Because feminism is the more likely possibility, Busemeyer et al. (2011) assumed that participants evaluate the conjunction as $P(F \text{ and then } BT)$, instead of in the

other order, which involves projecting the state vector, first onto the feminism subspace and then (without normalizing) onto the bank teller one. In Figure 1b, it can be seen that $P(F \text{ and then } BT) > P(BT)$.

More formally, the Linda story generates an initial state $|\psi_L\rangle$ in an N dimensional Hilbert space, the projector P_F is used to map the state onto the subspace for feminist, and the projector P_B is used to map the state onto the subspace for bank teller. Then the probability of the conjunction is computed by the quantum expression $P(F \text{ and then } BT) = \|P_B \cdot P_F \cdot |\psi_L\rangle\|^2$ and the marginal probability of bank teller equals $P(B) = \|P_B \cdot |\psi\rangle_L\|^2$. An order effect occurs when the two projectors do not commute, $P_F \cdot P_B \neq P_B \cdot P_F$, in which case the measurements are called incompatible.

The geometric character of quantum probabilities may tempt an inference that this is the main difference relative to Bayesian probabilities – and so perhaps it might suffice to label the general approach as just "projection geometry" (M. D. Lee personal communication, September 2023). However, quantum cognitive models also employ various key results from quantum theory about the way projections to subspaces correspond to probabilities and the interpretation of linear mixtures, called superpositions. We briefly mention four such results. First, quantum cognitive models follow Born's rule that probabilities are computed from squared magnitude of quantum states, after projections. Second, the models obey the remarkable Gleason's theorem showing that the quantum rule for associating probabilities to subspaces is the only possible way for doing so. Third, the models obey Kochen-Specker theorem which states that systems in superposition do not have definitive values until measured. Finally, the models follow Luder's law which determines how a state should update post-measurement. Note, Luder's law is the quantum equivalent of Bayes rule. It is the use of these key results which make more suitable the label 'quantum', or more precisely quantum-like, for these kind of models. Other work in psychology has employed projections (e.g., Sloman, 1993), but they did not use these additional theorems from quantum theory.

In previous work, we emphasized presentation in terms of vectors and projectors, as we wanted to explain the geometrical nature of quantum probabilities. In the current work,

the emphasis is on the probability relations derived from the geometrical properties of the quantum models, expressed in terms of the quantum interference term (Busemeyer et al., 2011). Accordingly, we will present the Quantum Sequential Sampler model using classical probabilities along with a quantum interference parameter, which can be turned on and off, to allow deviation and consistency with classical probabilities respectively. This will also help illustrate the Quantum Sequential Sampler as a hybrid model, encompassing both Bayesian and quantum probabilities.

Despite the difference in presentation, quantum probabilities in the Quantum Sequential Sampler can be seen as almost equivalent to that in the Busemeyer et al.'s (2011) model, except for one important difference when the quantum probabilities in the Quantum Sequential Sampler are right at the bounds. Busemeyer et al.'s (2011) model assumes projectors and as a result the Quantum Question (QQ) equality must be satisfied. However, when computing probabilities at the bounds, the Quantum Sequential Sampler model allows for violations of the QQ equality and positive-operator valued measures (POVMs) are employed, instead of projectors. The difference between a POVM and a projector is that for the former there is a small probability for a mismatch between measurement and projection. For example, in Figure 1b, an observer may decide that Linda is a feminist, but the mental state might accidentally project to the $\neg F$ subspace. Therefore, POVMs offer a mechanism for noise in probability calculations. A theoretical reason for employing POVMs is they are the appropriate approximations to projectors (Nielsen & Chuang, 2010), when describing processes in a subspace of a given Hilbert space, denoted as H. That is, if we assume that our knowledge is represented by a space of huge dimensionality H and a particular thought process requires focus/ restriction to a certain subspace, then projectors in H are approximated as POVMs in the subspace; this is Naimark's dilation theorem, e.g., Paulsen (2002). POVMs have been employed in some quantum cognitive models (White et al., 2020; Yearsley and Pothos, 2016, Lebedev & Khrennikov, 2024). We will return to this difference when we introduce the bounds formally.

The first steps in the quantum calculus for computing subjective probabilities are

essentially Bayesian, requiring us to specify three probabilities $P(A), P(B), P(B|A)$ for each pair of arbitrary events A and B. Like in the Bayesian Sampler, these are treated as free parameters in the model. Given these free parameters, we can then compute

$$\begin{aligned} P(\neg A) = 1 - P(A), P(\neg B) = 1 - P(B), \\ P(\neg B|A) = 1 - P(B|A). \end{aligned} \tag{9}$$

Equation (9) can be straightforwardly employed to compute two conjunctions and, using quantum interference/ order effect parameters o_1, o_2, we can compute the same conjunctions, but in the opposite order:

$$\begin{aligned} P(A \text{ and then } B) = P(A)P(B|A), \\ P(A \text{ and then } \neg B) = P(A)P(\neg B|A), \\ P(B \text{ and then } A) = P(A \text{ and then } B) - o_1, \\ P(\neg B \text{ and then } A) = P(A \text{ and then } \neg B) - o_2. \end{aligned} \tag{10}$$

With the aid of a third order effect parameter, o_3, we finally compute:

$$\begin{aligned} P(B \text{ and then } \neg A) = P(B) - P(B \text{ and then } A), \\ P(\neg B \text{ and then } \neg A) = P(\neg B) - P(\neg B \text{ and then } A), \\ P(\neg A \text{ and then } \neg B) = P(\neg B \text{ and then } \neg A) - o_3, \\ P(\neg A \text{ and then } B) = P(\neg A) - P(\neg A \text{ and then } \neg B). \end{aligned} \tag{11}$$

In general, the three interference effect parameters are bounded in the following way:

$$\begin{aligned} P(A \text{ and then } B) - P(B) \leq o_1 \leq P(A \text{ and then } B), \\ P(A \text{ and then } \neg B) - P(\neg B) \leq o_2 \leq P(A \text{ and then } \neg B), \\ P(\neg B \text{ and then } \neg A) - P(\neg A) \leq o_3 \leq P(\neg B \text{ and then } \neg A). \end{aligned} \tag{12}$$

The order effects are at the heart of a quantum probability model and quantify the extent to which $P(A \text{ and then } B) \neq P(B \text{ and then } A)$ (analogously for disjunctions). Thus, an ambiguity arises, concerning the way an observer would interpret the conjunction between A and B. Busemeyer et al. (2011, 2015) suggested that, unless primed in a specific way, observers process a conjunction in the order of the most likely predicate first. This is a necessary assumption for the emergence of conjunction fallacies: assume $P(A) > P(B)$.

Quantum theory is constrained so that $P(A \text{ and then } B) = P(A)P(B|A) \leq P(A)$. Therefore, we can only have conjunction fallacies of the form $P(B) \leq P(A \text{ and then } B)$, that is, for the less likely predicate. Busemeyer et al. (2011, 2015) further justified the 'more likely first' assumption by invoking the ideas in Gigerenzer and Goldstein (1996), concerning the prioritization of information. Concerning disjunctions, in quantum theory $P(X \text{ or then } Y) = 1 - P(\neg X \text{ and then } \neg Y)$, as is the case in Bayesian theory, but expressed in an order-specific way. Using the more likely first rule for $P(\neg X \text{ and then } \neg Y)$, the required order would be as shown, if $P(\neg X) > P(\neg Y)$. Therefore, following from the above example where $P(A) > P(B)$, the disjunction order would be $P(B \text{ or then } A)$. Noting that in quantum theory $P(B \text{ or then } A) \geq P(B)$ the only allowed disjunction fallacy would be of the form $P(B \text{ or then } A) < P(A)$, that is, in relation to the more likely predicate, as expected.

When computing the conditionals, the order of the conjunctions still matters, and they are computed as $P(X|Y) = \frac{P(Y \text{ and then } X)}{P(Y)}$, for arbitrary events X and Y. Given the assumption that the more likely event is always processed first, one might question the need for the interference term at all. However, this is needed for the computations involving the marginal probability of the less likely event and the conditional probability of the more likely event, given the less likely event. For example, $P(B) = P(B \text{ and then } A) + P(B \text{ and then } \neg A)$, and $P(A|B) = \frac{P(B \text{ and then } A)}{P(A)}$.

Accordingly, the quantum model requires, at most, six parameters for each pair of questions, $\{P(A), P(B), P(B|A), o_1, o_2, o_3\}$; note, all classical probabilities can be computed using just three parameters. For the three pairs of questions we explored empirically, the potential number of parameters grows to 15. Note also that o_1, o_2, o_3 may have different bounds for different pairs according to Equation 12. Initially, assume that all of o_1, o_2, o_3 are in the bounds of each other. To reduce modeling complexity, we adopted the following assumptions.

First, following Busemeyer et al. (2011), we assume initially that the measurements are performed by projectors. This implies that the QQ equality is satisfied (Busemeyer

et al, 2011), which is equivalent to assuming that $o_1 = o_3$. As noted, quantum theory includes more general measurement operators, POVMs, which do not necessarily satisfy the QQ equality, in which case o_1 is not required to equal o_3 (Yearsley and Busemeyer, 2016). However, to start with, we restrict the model to satisfy the QQ equality when the bounds are not violated.

Second, we assume that the interference effects, o_1, o_2, o_3, are the same across the three pairs of questions. This is reasonable because each pair of questions is about a pair of state election results for the same candidates. This is analogous to the assumption in the Bayesian sampler that the sample size is constant across pairs.

Third, we equated interference effects o_1 and o_2 as follows: Suppose $P(A) \geq P(B)$, we assume that $o_2 = -o_1$; suppose $P(A) < P(B)$, we assume $o_2 = o_1$. These assumptions imply that a conjunction error can only occur with the less likely event when $o_2 \neq 0$. More formally,

$$\begin{aligned} P(B) &= P(B \text{ and then } A) + P(B \text{ and then } \neg A) \\ &= P(A \text{ and then } B) - o_1 + P(\neg A \text{ and then } B) + o_2 \\ &= P(A \text{ and then } B) + P(\neg A \text{ and then } B) - (o_2 - o_1). \end{aligned} \tag{13}$$

A similar derivation shows that

$$P(A) = P(B \text{ and then } A) + P(\neg B \text{ and then } A) - (o_2 + o_1). \tag{14}$$

The interference $-(o_2 + o_1)$ is zero when $o_2 = -o_1$ and non-zero when $o_2 = o_1$, and vice versa for the term $-(o_2 - o_1)$. Thus, by equating o_2 and o_1 this way, the model produces interference only for the less likely event. Note that we need $-(o_2 - o_1)$ to be negative to produce $P(B) < P(A \text{ and then } B)$ when $P(A) > P(B)$, and vice versa when $P(B) > P(A)$. Therefore, having interference on the less likely event biases the model to identify conjunction fallacies, remembering that in the quantum model conjunction fallacies arise only against the less likely predicate.

However, as mentioned, o_1, o_2, o_3 are bounded differently for each pair. Therefore, when one of the interference effect parameters is greater than the upper bound or smaller

than the lower bound of the other interference effect parameter, it is not possible to set $o_1 = o_2 = o_3$ or $o_1 = -o_2 = o_3$. To circumvent this problem, we adopted an additional assumption that the interference effect parameters whose bounds are violated by other interference effect parameters would be set to the values of their bounds being violated **(the bound assumption)**. To illustrate, consider the case when $P(\neg B \text{ and then } \neg A) - P(\neg A) < o_3 < P(A \text{ and then } B) - P(B)$, and both o_1 and o_3 are within the bounds of o_2. In this case, since o_3 is less than the lower bound of o_1, it is not possible to set $o_1 = o_3$. Under the bound assumption, we would instead reset $o_1' = P(A \text{ and then } B) - P(B)$ so that it is as close to o_3 as possible, without violating the axioms of quantum probabilities. Similarly, when o_2 is initially set to $-o_1$, but $-o_1 > P(A \text{ and then } \neg B)$, we would reset $o_2' = P(A \text{ and then } \neg B)$.

Following the bound assumption, there are two consequences when the bounds of the interference effect parameters are violated by the other interference effect parameters. First, since the interference effect parameters are reset to their bounds when violated, the QQ equality may no longer hold for certain parameter values. This is evident in the previous example, where o_1 is reset to $o_1' = P(A \text{ and then } B) - P(B)$ and $o_3 < P(A \text{ and then } B) - P(B)$. However, as mentioned, a violation of QQ equality is possible in quantum probability when POVMs are employed. Second, it is now possible that there could be a non-zero interference effect for the more likely predicate, even though there would still be no conjunction error for the more likely predicate. Consider the example where $P(A) > P(B)$. According to Equation 14, the interference term of the more likely predicate is $-(o_2 + o_1)$. When no bound is violated, this interference term is set to zero. On the other hand, when the lower bound of o_2 is violated by $-o_1$, we set $o_2 = P(A \text{ and then } \neg B) - P(\neg B) < -o_1$, and thus $-(o_2 + o_1) \neq 0$. However, there must still be a non-zero interference effect for the less likely event with the bound assumption. To see why, suppose $-o_1$ violates the upper-bound of o_2 and o_2 is initially set to $-o_1$. Then, since $-o_1 > P(A \text{ and then } \neg B) \geq 0$, it must be the case that $o_1 < o_2 = P(A \text{ and then } \neg B)$ and thus $o_2 - o_1 \neq 0$. Conversely, for $-o_1$ violating the lower bound of o_2. The same logic applies to when o_2 is initialized as o_1. Therefore, the bound assumption does not alter the

quantum model's ability to produce conjunction fallacy for the less likely predicate.

To summarize, we computed the quantum probability part of the model through the following procedure: we first assume that $o_1 = o_3$ and set $o_2 = -o_1$when $P(A) \geq P(B)$, and assume that $o_2 = o_1$ when $P(A) < P(B)$. Next, we checked the bounds of the three interference effect parameters, using the initialized values of these order effect parameters. We then constrained the interference effect parameters to equal the violated bounds, if any bound were violated. Finally, we used the constrained interference effect values, the Bayesian probabilities, and the more likely first assumption to compute all quantum probabilities.

The key feature of this formulation of the quantum probability part is that it allows seamless transition between strongly quantum probabilities and strongly Bayesian ones, simply by virtue of the size and necessity of the interference effect parameter. At a preliminary level, we can ask whether the model where $o_1, o_2, o_3 \neq 0$ is needed at all versus when all of the order effects parameters are 0. Note the Bayesian (classical) variant also requires that $P(B) > P(A \text{ and then } B)$, $P(\neg B) > P(A \text{ and then } \neg B)$, and $P(\neg A) > P(\neg B \text{ and then } \neg A)$ are satisfied. The Bayesian and the quantum interference variants of the model are nested. Comparing these two model versions with a G^2 test over all participants revealed that 576 out of 1162 participants were better fitted by the version with non-zero quantum interference parameters. Given that the percentage of participants better fitted by the quantum interference version is 50%, much higher than expected by chance using a 5% significance level, we will use the quantum interference variant of the model in further analysis. Recall that this is much higher than 0.6% of significant improvement of the more complex version of the Bayesian Sampler from the simpler version. The two variants of the models were also compared using a generalization test, where the quantum variant outperformed the classical variant for both tests. The details of the generalization test will be discussed in the Model Comparisons Section.

Quantum Sequential Sampler second part – sequential sampling process . We assume that subjective probabilities cannot be used for responding directly, but rather correspond to drift rates in a sequential sampling process, which eventually results in probability responses.

There is considerable evidence that probability judgments are not just a simple linear transformations of subjective probabilities (see Wallsten & Budescu, 1983, for a review). Even if subjective probabilities are mentally represented, they are likely to be unconscious, uncertain (e.g., because of concerns with the precision or fidelity of information), and lack clarity. Therefore, a sampling estimation process is still required to convert what could be vague information concerning probabilities to actual probability estimates, that is, probability ratings. Additionally, there might be other reasons a probability response might be a distorted version of a subjective probability. For example, a cognitive agent may feel they are unable to accurately estimate subjective probabilities (e.g., because of time pressure) or is intentionally seeking to distort subjective probabilities (e.g., because probability estimates biased in a certain direction serve a particular purpose). Such considerations justify the assumption that subjective probabilities are best approached as drivers of a response process, rather than directly corresponding to responses themselves.

Costello and Watts (2014) and Zhu et al. (2020) pioneered the idea that sampling can be employed as the response process. However, as mentioned, these sampling models require a commitment to sample size prior to any evidence about the relevant probabilities and independently of any dynamic task demands, such as a prompt to hurry up with a judgment, after the start of the judgment process. An alternative proposal is that the response process is a sequential sampling one, where evidence is gradually accumulated towards the available responses, until a stopping criterion is reached (Ratliff & Smith, 2015). There is extensive experimental evidence for sequential sampling processes (e.g., Brown & Heathcote, 2008; Diederich, 2003; Johnson & Busemeyer, 2005; Trueblood et al., 2014; Usher & McClelland, 2001), as well as neuroscience evidence of such processes in the brain. Including a sequential sampling component to our model extends the predictive scope to encompass reaction times, confidence ratings, and uncertainty in choice behavior (e.g., Busemeyer & Diederich, 2009; Ratcliff, 1978; Ratliff & Smith, 2015; Ratcliff et al., 2016; Usher & McClelland. 2004), which could be exploited in future extensions of paradigms for probabilistic reasoning. As a technical point, a sequential sampling process offers prediction across the range of possible

ratings.

Despite being inspired by traditional sequential sampling models for evidence accumulation processes, the sequential sampling part of Quantum Sequential Sampler has one key difference when compared to them. Typically, states in evidence accumulation models represent evidence that cannot be directly measured in experiments, with evidence accumulating towards a specific boundary for choice responses. In contrast, the states in Quantum Sequential Sampler model are directly measurable responses. The model operates within a vector space where states denote probability judgments, making even intermediate states measurable and interpretable.

The ability to measure intermediate states as probability judgments enables the Quantum Sequential Sampler to have two different interpretations for different experimental tasks. First, for probability judgments, which is the primary concern of this paper, the model performs a continuous update of the probability judgment distribution with a Markov process in light of evidence from mental simulations. In such a case, the distribution of the probability judgments evolves deterministically following the Kolmogorov equation. However, when one obtains a sample point from the probability distribution, such a sample point is obtained randomly from the distribution. The uncertainty in our model is meant to correspond to people's assumed inherent uncertainty about the exact value of a probability judgment for a particular event. In fact, this is very similar to Bayesian belief updating, but instead we use a continuous time Markov process for a continuous time update. That is, the previous state at time t can be seen as a Bayesian prior and the next state in time $t + \Delta t$ can be seen as a posterior, after some evidence has accumulated. The model stops after running for a fixed duration, determined by stopping condition determined by working memory capacity and cognitive loads (Usher & McClelland 2001; Ratcliff, 2006).

Second, the Quantum Sequential Sampler could be applied for measuring choice and response time. In this case, the Quantum Sequential Sampler functions exactly the same as traditional evidence accumulation models except that the evidence states are interpretable. As in traditional evidence accumulation models, stochasticity in the sequential sampling

part now arises from noisy updating of the state at each time step, until it hits one of the boundaries. The choice probability is then computed as the proportion of hits. Since we are concerned with probability judgment instead of choice and response time, we focus on the first interpretation in the present work.

Formally, the sequential sampling part can be specified using a discrete state Markov process or a continuous state diffusion process. In most practical cases, probability judgments are expressed as integers on some scale. Even when using an approximately continuous scale for ratings, positions on the scale are actually discrete. Therefore, below we present the corresponding discrete state Markov process (cf., Busemeyer et al., 2006; Appendix 1, describes the corresponding continuous state diffusion process).

We start by assuming $N = 101$ states representing probability ratings on an integer scale from $i = 0, 1, 2, ..., 100$. Before evaluating the probability judgment, the person starts with an $N \times 1$ initial state vector $\phi(0)$, which is a probability distribution across the states that sums to unity. The coordinate, $\phi_i(0)$, is the probability of starting at probability judgment i. We define this initial state $\phi(0)$ with a symmetric beta distribution $Beta(\gamma, \gamma)$, which is the same Bayesian prior as that employed in Zhu et al. (2020) – the same initial condition is used for the diffusion model in Appendix 1. Since the beta distribution is a continuous-space distribution, we need to discretize it for the Markov model with N states and we do so by using the same technique as for the Bayesian Sampler model (Equation 9). Also, given that $\phi(0)$ represents a probability mass function at time 0, we normalize the density mapping u in Equation (7).

During the evaluation process, the distribution across states evolves according to a Markov process with a drift rate determined by the subjective probability obtained from the quantum probability model. The evolution of the Markov process is determined by the following Kolmogorov forward equation

$$\frac{d}{dt}\phi(t) = K \cdot \phi(t), \tag{15}$$

which has the solution

$$\phi(t) = e^{K \cdot t}\phi(0). \tag{16}$$

In the above, K is an $N \times N$ intensity matrix which encodes the state transition rates, $\phi(t)$ is a time-dependent $N \times 1$ vector that encodes the probability mass function across the N states (where each state represents a probability judgment value in this case), and $\phi(0)$ is the initial probability mass function. To define the Markov process that maps the subjective probabilities to probability judgment responses, we therefore need to define the intensity matrix K.

For each subjective probability $P(A)$, where the event A can be a marginal, conjunction, disjunction, or conditional, the intensity matrix K with a reflecting boundary can be specified as follows

$$
\begin{aligned}
&K_{i,i+1} = \beta_+ \text{ for } 1 \leq i \leq N-1 \\
&K_{i+1,i} = \beta_- \text{ for } 1 \leq i \leq N-1 \\
&K_{i,i} = -(\beta_+ + \beta_-) \text{ for } 2 \leq i \leq N-1 \\
&K_{1,1} = -\beta_+ \\
&K_{N,N} = -\beta_-,
\end{aligned}
$$

$$
K = \begin{bmatrix}
-\beta_+ & \beta_- & 0 & \cdots & 0 & . \\
\beta_+ & -(\beta_- + \beta_+) & \beta_- & \cdots & . & . \\
0 & \beta_+ & -(\beta_- + \beta_+) & \cdots & . & . \\
 & 0 & \beta_+ & \cdots & . & . \\
. & . & 0 & \cdots & . & . \\
. & . & . & \cdots & 0 & . \\
. & . & . & \cdots & \beta_- & . \\
. & . & 0 & \cdots & -(\beta_- + \beta_+) & \beta_- \\
0 & 0 & 0 & \cdots & \beta_+ & -\beta_-
\end{bmatrix}. \tag{17}
$$

The parameters β_-, β_+ are also employed in the diffusion model (Equation A3.2) and are given by

$$
\begin{aligned}
\beta_+ &= \alpha \cdot P(A) + c_+ \\
\beta_- &= \alpha \cdot (1 - P(A)) + c_-,
\end{aligned} \tag{18}
$$

where $\alpha \geq 0$ moderates the effect of the subjective probability on the drift parameter. The constants c_+, c_- are further defined by a free additive bias parameter b as follows. If $b > 0$ then $c_+ = 1 + b$ and $c_- = 1$; if $b < 0$ then $c_+ = 1$ and $c_- = 1 - b$; finally if $b = 0$ then $c_+ = c_- = 1$. These assignments guarantee that the intensity matrix parameters are always

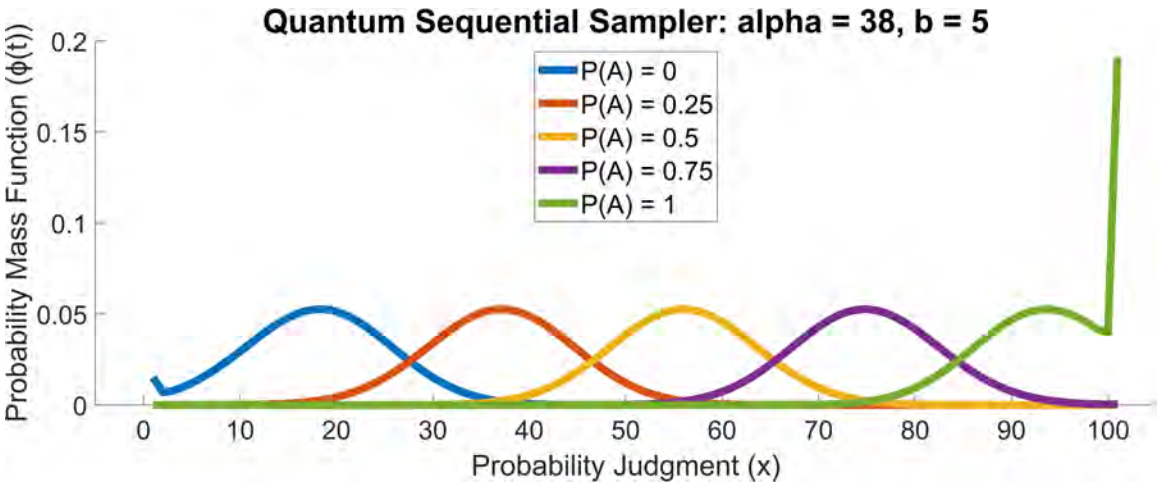


Figure 5. An illustration of the sequential sampling part of the model. The α parameter for the above model is 38 and the bias parameter b is 5. As can be seen, with a positive bias parameter, $\phi(0)$ drifts faster towards the right, to produce probability judgments greater than 0.5, than it drifts towards the left, to produce probability judgments less than 0.5.

positive.

Conceptually, β_+ and β_- represent the transition rate of increase and decrease, respectively, in a probability judgment over time. Their difference determines the drift rate and their sum determines the diffusion rate (see Appendix 1 for how drift and diffusion rates influences the change in means and variances of the Markov process). Parameters α, c_+, c_-, for specifying β_+ and β_-, control the strength of transition rates. Note that since c_+, c_- are defined through b, we only need two free parameters, α and b, along with the relevant subjective probability, to specify the intensity matrix for each probability question; we assume α and b to be the same for different probability questions.

Behaviorally, the quantities β_+, β_- embody two key mechanisms. First, they depend on the subjective probabilities for a particular question, thus linking responses to a veridical, internal probabilistic process by the agent. Second, they embody biases which allow over- (or under-) estimation of probabilities and so systematic biases in probability estimation. Note, this bias is blind to the probability question at hand, that is, it does not a priori differentiate between conjunctions and marginals, which contrasts with the corresponding key parameters in Costello and Watts (2014) and Zhu et al. (2020). Specifically, when $b > 0$

versus $b < 0$ people overestimate versus underestimate probability judgments, relative to the subjective probabilities (Figure 5). That is, $b > 0$ could be interpreted as an overestimation bias, which exists prior to any mental simulation, while $P(A) \cdot \alpha$ and $(1 - P(A)) \cdot \alpha$ reflect the evidence gathered from mental simulations, which regulates this preexisting bias.

To fit the Quantum Sequential Sampler model to data, we need to define a likelihood function. Let x be an observed probability judgment and denote the $x + 1$ element of the final state vector (assuming the first index is 1) at time t as $\phi_(x+1)(t)$. Then the likelihood of the judgment at response time t can then be written as:

$$L(x, t|model) = \phi_{x+1}(t). \tag{19}$$

For example, for probability judgments corresponding to integers from 0 to 100, if we have $x = 50$, then $L(x = 50, t|model) = \phi_{51}(t)$. That is, the prediction of the model when observing a probability judgment of 50 would be the 51st entry of $\phi(t)$. Note that for each event A, there will be one corresponding likelihood in equation 19 and the final aggregated likelihood will be the product of each of these likelihoods. While this Markov process allows for time dependence, our current data do not track time and so fits proceeded assuming that the response time is the same for all judgments. This means that the time parameter can be absorbed into the other parameters of the intensity matrix; time is just a constant multiplying to the matrix. However, the Quantum Sequential Sampler can be extended to allow time to vary according to experimentally measured response times – this is just a matter of de-clamping time from existing parameters. In the future, it would be worthwhile to manipulate time, with a view to examine whether the present model can jointly predict response time and probability judgments.

There is an additional remark regarding the fixed response time in our model as compared to the fixed sample size in the Bayesian Sampler model. We do not make any assumption that a fixed response time has to map to a single fixed sample size and in fact it is possible that varying degrees of response time can correspond to the same sample size. In other words, even if response time can vary, our model does not contradict the assumption that sample size might be fixed. Besides, we also do not assume a process of how samples

are drawn: samples could be autocorrelated or drawn in parallel or maybe what is needed is even a partial simulation sample (Bass et al., 2022). To sum up, despite the fact that sample size is related to response time, it has no functional role in how the Quantum Sequential Sampler produces predictions. This contrasts with the Bayesian Sampler model: in the Bayesian Sampler, sample size is a parameter which has to be fitted directly as part of explaining a set of probability judgments.

In conclusion, we followed mostly standard formalism for Markov processes, with a particular definition for the intensity matrix. The intensity matrix is standard for any Markov version of a random walk model (e.g., Busemeyer & Diederich, 2009; Ratcliff & Smith, 2015). Together with the subjective probability part, the Quantum Sequential Sampler has the following ten parameters $\{P(A),\ P(B),\ P(C),\ P(B|A),\ P(C|A),\ P(C|B),\ o,\ \gamma,\ \alpha,\ b\}$, where o is the interference parameter, γ determines the initial distribution across ratings, and α, b are used to determine the drift rates of the Markov model. Overall, while sequential sampling processes have been widely employed in judgment and decision making, to the best of our knowledge this is the first time they are applied to probabilistic reasoning.

Analytical predictions: Binary Complementarity. The question of whether analytical mean predictions can be derived from the Quantum Sequential Sampler model, analogous to those for the Bayesian Sampler model, presents an intriguing line of inquiry. However, this task presents substantial challenges, largely due to the intricate nature of the Markovian dynamics and the implementation of reflecting boundaries. Despite these difficulties, linear approximation is an effective method to acquire some insights into model behavior and predictions. In Appendix 3, we examine analytically model behavior for various major probabilistic fallacies, including conjunction and disjunction fallacies, as well as violations of probability identities.

Here, we show how linear approximation can illuminate the ways in which the Quantum Sequential Sampler model addresses violations of binary complementarity, a fallacy that is prominently represented in our current dataset. Amongst the several kinds of probabilistic fallacies in our data (and previous work), binary complementarity uniquely distinguishes

between the Quantum Sequential Sampler and the other computational models: as far as we know, the Quantum Sequential Sampler is the only model which can account for violations of binary complementarity. In this section, we explain how.

For a Markov process characterized by a constant intensity matrix and initialized from a symmetric beta distribution, the mean of an arbitrary event A is anticipated to exhibit a roughly linear increment with time, adhering to the relationship according to Equation A.1.7 (see Appendix 1 for more details):

$$\mu_{QSS}(t, A) \approx \frac{1}{2} + (\beta_+ - \beta_-)t, \tag{20}$$

where $\mu_0 = \frac{1}{2}$ represents the mean of the symmetric beta distribution and $(\beta_+ - \beta_-)$ delineates the drift rate inherent to the Markov process. While $\mu_{QSS}(t, A)$ represents probability judgments on a scale from 0 to 1, note that in the current dataset judgments were fitted as integers from 0 to 100. The [0, 1] scale is employed here to maintain consistency with the approach used by Costello and Watts (2014) in their demonstration of violations of probability identities.

By incorporating the expressions for β_+ and β_- from Equation 19, we arrive at:

$$\mu_{QSS}(t, A) \approx \frac{1}{2} + \{\alpha(2P(A) - 1) + b\}\, t = \frac{1}{2} + 2\alpha t P(A) + (b - \alpha)t. \tag{21}$$

Under the assumption that the processing time t remains constant for any event A, it becomes feasible to absorb t as a constant factor into the parameters d and k. As noted, this modeling assumption is specific to the dataset at hand, but the model retains its functionality even when t varies across judgments. This results in the simplified expression:

$$\mu_{QSS}(A) \approx \frac{1}{2} + 2\alpha P(A) + (b - \alpha). \tag{22}$$

Using the above expression, the Quantum Sequential Sampler can predict violation of binary complementarity as follows: for any event A

$$\begin{aligned} \mu_{QSS}(A) + \mu_{QSS}(\neg A) &\approx \frac{1}{2} + 2\alpha P(A) + (b - \alpha) + \frac{1}{2} + 2\alpha P(\neg A) + (b - \alpha) \\ &= 1 + 2\alpha(P(A) + P(\neg A)) + 2(b - \alpha) \\ &= 1 + 2b, \end{aligned} \tag{23}$$

where α is the drift rate parameter, b is the overestimation/underestimation parameter, and $P(A) + P(\neg A) = 1$. The same result holds for conjunctions and disjunctions as well, since in the Quantum Sequential Sampler model, regardless of whether A or B is measured first, it must be the case that $P(A \wedge B) + P(\neg A \vee \neg B) = 1$.

From Equation 23, it is clear that the sum of the judgments for an event and its negation can deviate from 1, depending on the value of b:

- When $b > 0$, we observe a subadditivity effect, resulting in an overestimation.
- When $b < 0$, we observe a superadditivity effect, resulting in an underestimation.

In summary, the sequential part of the Quantum Sequential Sampler can be distorted based on parameter b, allowing the model to explain phenomena beyond the reach of the Bayesian Sampler model. In the present study, this translates to a general overestimation of probabilities, corresponding to a positive value of b. While we think this analysis illustrates reasonably well one source of advantage for the Quantum Sequential Sampler, there are two important caveats. First, actual model behavior is more complex than the linear approximation warrants. Exploring full model behavior cannot be done analytically and simulation methods would be the only viable approach. Second, the capacity of the Quantum Sequential Sampler to explain violations of binary complementarity is a theoretical advantage. However, this is not the only reason for the model's advantage over other models. For example, while the Bayesian Sampler model cannot account for conjunction fallacies when both marginals and conjunctions are greater than 0.5, our model can explain such fallacies, because of quantum probabilities. In general, the predictions from the Quantum Sequential Sampler depend on both the sequential sampling part and the quantum probabilistic calculus.

Model comparisons

We perform our model comparisons at the individual level using log likelihood criteria, and explore its ability to capture all the probability judgments in the present dataset. This approach subsumes all individual fallacies and effects, which have previously been discussed

in the literature, including some new effects we identified. Two different log likelihood criteria are employed. The first is to estimate the parameters using all 78 ratings from each participant and compare models using the Bayesian Information Criterion, BIC $= -2 \cdot G^2 + p \cdot ln(N)$, where p = number of parameters and $N = 78$ observations per person. The Bayesian Information Criterion (BIC) serves as an approximate measure for the Bayes factor. Under ideal circumstances, the Savage-Dickey method would be the preferred choice for accurately determining the Bayes Factor (Lee & Wagenmakers, 2014), but this approach is too computationally expensive for our models. Despite facing criticism, the BIC remains a popular tool for model comparison within the field of psychology. To enhance our analysis beyond the limitations of BIC, we have also employed a generalization test method, which provides a more rigorous evaluation of model complexity (Busemeyer and Wang, 2000). This involves estimating model parameters using a calibration set of conditions and subsequently applying these parameters to predict outcomes for a generalization set of conditions. Compared with cross-validation methods often used inmachine learning models (LeCun et al., 1998; Lecun, Bengio, & Hinton, 2015; Nair & Hinton, 2010), generalization tests are similar but even more rigorous. In the generalization test we employed, instead of merely excluding random data points to form a test set, we strategically removed key elements, such as conjunctions and disjunctions, thereby subjecting the model to a more rigorous evaluation.

Regarding models to compare, we can only compare models that can make predictions for all 78 probability queries. Reviewing the points made above, some models, like the averaging model (Nilsson et al., 2009) or inductive confirmation (Tentori et al., 2013), have not been formulated in a way that allows predictions for some queries, such as conditional probabilities. Other models, like the probability-plus-noise model (Costello & Watts, 2014), which use a binomial distribution, produce zero likelihoods for some ratings, which make comparisons based on log likelihood impossible. Furthermore, the Bayesian Sampler is a similar and arguably better account compared to the probability-plus-noise model, as shown in Zhu et al. (2020). Finally, with such a large number of participants to fit, it is very costly in time to fit many models. Therefore, we focus on comparing four models: simple and

complex forms of the Bayesian and quantum sequential samplers.

Regarding the comparisons between simple and complex versions of each model, we remind readers that these comparisons were discussed earlier in the article. In the previous section presenting the Bayesian Sampler, we summarized a comparison of a simpler version, using only two sample sizes, with a more complex model, using four sample sizes. The statistical results reported in that section did not indicate a rejection of the simpler model and so we will use the simpler version that includes only two sample sizes for model comparisons. In the section presenting the Quantum Sequential Sampler, we summarized a comparison of a simpler version with no interference, which we refer to as the classical variant or Classical Sequential Sampler, to a more complex model which included an interference parameter, which we refer to as the quantum variant. Statistical results favored the quantum variant and it is this version which will be the focus for the comparison with the Bayesian Sampler.

BIC comparison

Both the Quantum Sequential Sampler and the Bayesian Sampler were fitted to individual participant judgments through maximizing log-likelihood, which can be converted to G^2 values. We also evaluated a baseline uniform distribution model that uniformly randomly guesses integers from 0 to 100 as probability judgments. The models were compared using the BIC score averaged across all participants, computed from mean G^2, with appropriate penalties for the number of parameters of the models. The baseline model (random rating) produced a BIC of 718.41, which is much higher than both models. The Quantum Sequential Sampler model (10 parameters) produced a mean BIC of 616.53, which is much lower than that of the Bayesian Sampler model (9 parameters) at 662.94. The classical variant of the Quantum sequential sampler model (9 parameters), which assumes no quantum interference, is also much better than the Bayesian Sampler model, with a BIC of 618.32. The classical and quantum variants of the Quantum Sequential Sampler have comparable BIC results, but the quantum variant performs slightly better.

We also used BIC to examine the number of participants fitted better by either model. Consistently with the above conclusion, the Quantum Sequential Sampler produces a lower

	QSS	CSS	BS
Conjunction Training	396.11	400.39	418.44
Conjunction Test	183.84	185.78	231.21
Disjunction Training	384.77	386.03	429.69
Disjunction Test	209.63	216.98	204.68

Table 3: Generalization test results (mean G^2 value) for the Quantum Sequential Sampler (QSS), the classical variant of Quantum Sequential Sampler (CSS), and Bayesian Sampler models (BS). 'Conjunction training' refers to the G^2 when training the model on all probabilities, apart from conjunctions and 'Conjunction test' when testing the trained model on conjunctions; analogously for disjunctions.

BIC value for 66% of all participants (769 counts). Therefore, according to mean BIC, the Quantum Sequential Sampler is a better model. Note, these statistics apply to testing both triplets; recall, we used two triplets of events for probability judgments, in a between participants condition. Hereafter, in most cases we report only aggregate results for both triplets for ease of presentation. Where aggregate results are reported, results broken down by triplet offer minor variations to the overall pattern, without altering conclusions.

Similar analyses were carried out for the Zhu et al. (2020) data set. The aggregated BIC over all five weather conditions for the Bayesian Sampler is 1607 and for the Quantum Sequential Sampler 1610, which shows that the two models have comparable performance (details in Appendix 2 and Supplementary Material 8). There are some possible reasons why Quantum Sequential Sampler did not achieve a clear advantage over the Bayesian Sampler in the data set of Zhu et al. (2020), as was the case in our dataset. Notably, responses in Zhu et al. (2020) were entered as numbers into a computer and judgments to the same events were measured repeatedly. With this response mode, it is possible that there is a stronger bias in reporting integers 5s and 10s (Budescu et al., 1988). Additionally, an earlier answer could bias subsequent responses for the same event, producing dependencies across replications. This in turn questions the suitability of a log likelihood approach that assumes independent observations. Another reason is that questions about the weather might be more likely to be represented in a compatible way, e.g., because of familiarity of such questions (Trueblood et al., 2017; Yearsley & Trueblood, 2018). Appendix 2 offers a more detailed discussion regarding the difference between the two datasets.

Generalization test

A stronger comparison of the models is obtained using a generalization test, which addresses the issue of model complexity in a more general manner (Busemeyer & Wang, 2000). We adopted two approaches: first, we trained each model on all the probabilities except conjunctions and then tested the model on conjunctions. Second, we trained each model on the all probabilities except disjunctions and then tested the model on disjunctions. This is a conservative test of the Quantum Sequential Sampler: when fitted across the entire set of judgments, the interference parameter balances the inflation of conjunctions and deflation of disjunctions. But considering each set of judgments individually removes this advantage.

The results of this generalization test are presented in Table 3, which shows the mean G^2 across all participants separately for the training set versus the test set and for conjunction versus disjunction test conditions. The Quantum Sequential Sampler performs better by a large margin for the conjunction test set, and the Bayesian sampler performs better by a small margin for the disjunction test set. The total G^2 across both test conditions for the generalization test set favors the Quantum Sequential Sampler (total G^2 equals 392) over the Bayesian sampler (total G^2 equals 436). Comparing the quantum variant of Quantum Sequential Sampler with its classical variant, the quantum variant outperforms the classical variant in both generalization tests.

We again examined the percentage of individuals for whom each model performed better. The results again support the Quantum Sequential Sampler model (quantum variant), with 79% of participants (921 counts) with a lower G^2 on the conjunctions test and 52% of participants (604 counts) with a lower G^2 on the disjunction test set. In sum, the Quantum Sequential Sampler overall outperforms the Bayesian Sampler, in terms of these generalization tests. Comparing the classical variant and the quantum variant of the Quantum Sequential Sampler with this generalization test, we also found that the quantum variant outperformed the classical variant, with 58% (679 counts) of participants better with a lower G^2 on the conjunctions test and 60% (702 counts) better on the disjunction test. The gen-

eralization test result further emphasizes the need of quantum probability, at least for some of the participants in our dataset.

Predictions

We compared the mean predictions to the mean response, across all participants, for each probability question, and the distribution of predictions compared to the distribution of responses, across all participants, again for each probability question. The prediction from the Quantum Sequential Sampler for an arbitrary probability question A is computed as the expected value of the final distribution of the Markov process, for arbitrary probability question A, which is:

$$Pred(A)_{QSS} = \sum_{i=o}^{100} i \cdot \phi_{i+1}(t)[P(A)] \tag{24}$$

The prediction for the Bayesian Sampler is computed in the same way as in Zhu et al. (2020):

$$Pred(A)_{BS} = 100 * \frac{NP(A) + \beta}{N + 2\beta} \tag{25}$$

Note, the '100' factor is used to convert probabilities to integers from 0 to 100, which correspond to the possible responses. So as to have a baseline model against which to compare the Quantum Sequential Sampler and the Bayesian Sampler, we also fitted the relative frequency model (also used in Zhu et al., 2020), which computes probability predictions based on relative percentages in a binomially distributed sample. The prediction of this relative frequency model corresponds to the relative percentage of the binomial mean, which is simply Bayesian probabilities, and thus is expected to strictly follow Bayesian probability axioms. Since the likelihood distribution of the relative frequency model is binomial, this model cannot be fitted with a likelihood-based method, and we thus fitted the model in the same way as Zhu et al. (2020), using sum of square error to compare the data against predictions from the relative frequency model. Note, this is a different baseline model from

[5] In Figure 6, we found a systematic overestimation bias so that for most judgments the mean is above 0.5. We checked and ensured that these are the correct results (See Figure S.2.2). Similar overestimation effects were found in Epping and Busemeyer (2023).

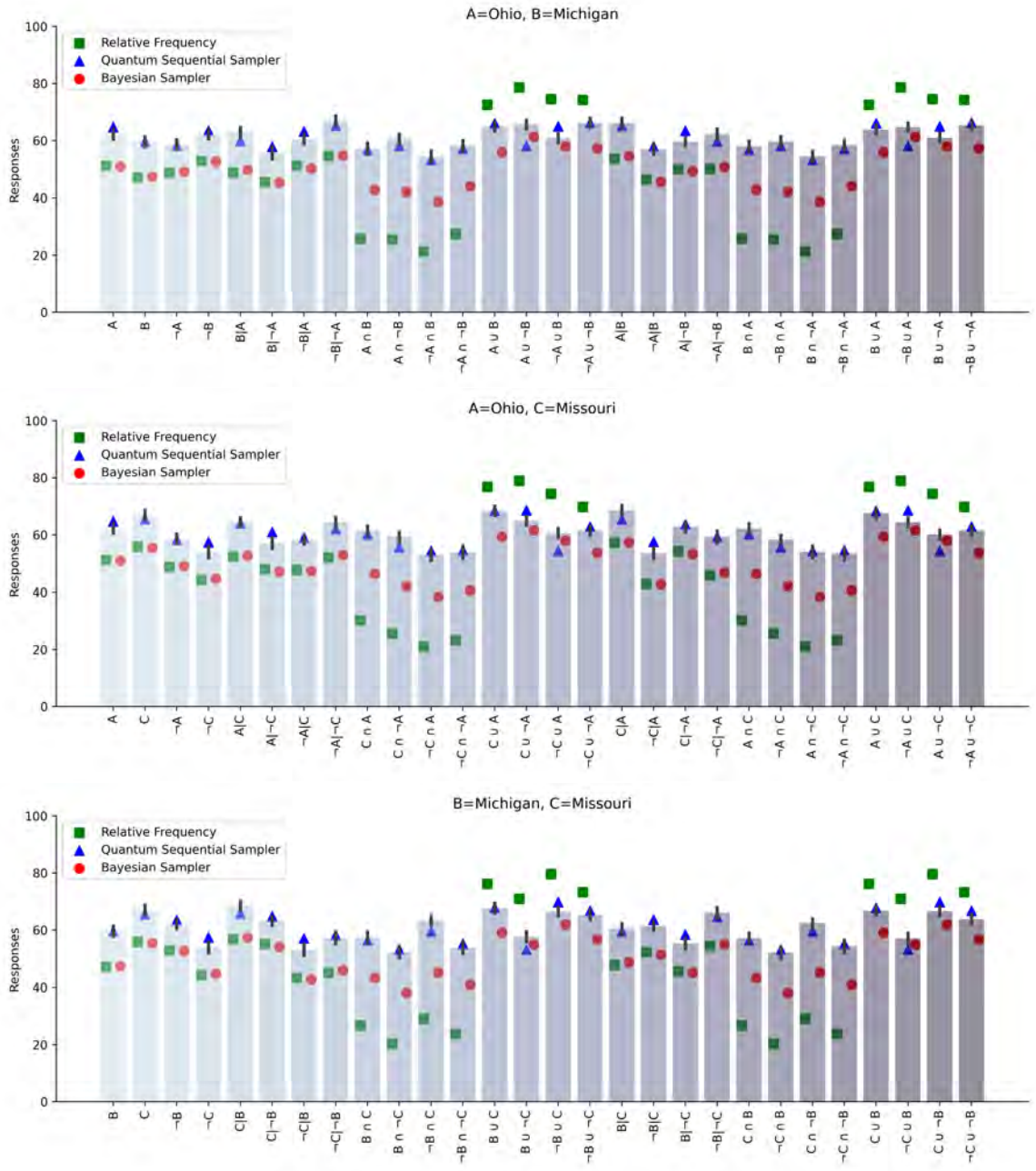


Figure 6. Mean predictions of the Quantum Sequential Sampler, Bayesian Sampler, and relative frequency models, against empirical results.[5] Here and elsewhere the responses on the vertical axis show probability judgments that the events on the horizontal axis occur on the scale [0,100]. Events A, B, C correspond to Biden winning the different states, as shown above, and negations correspond to Trump winning. The error bar in the middle shows the 95% confidence interval of the means. The results are for the $T1$ triplet.

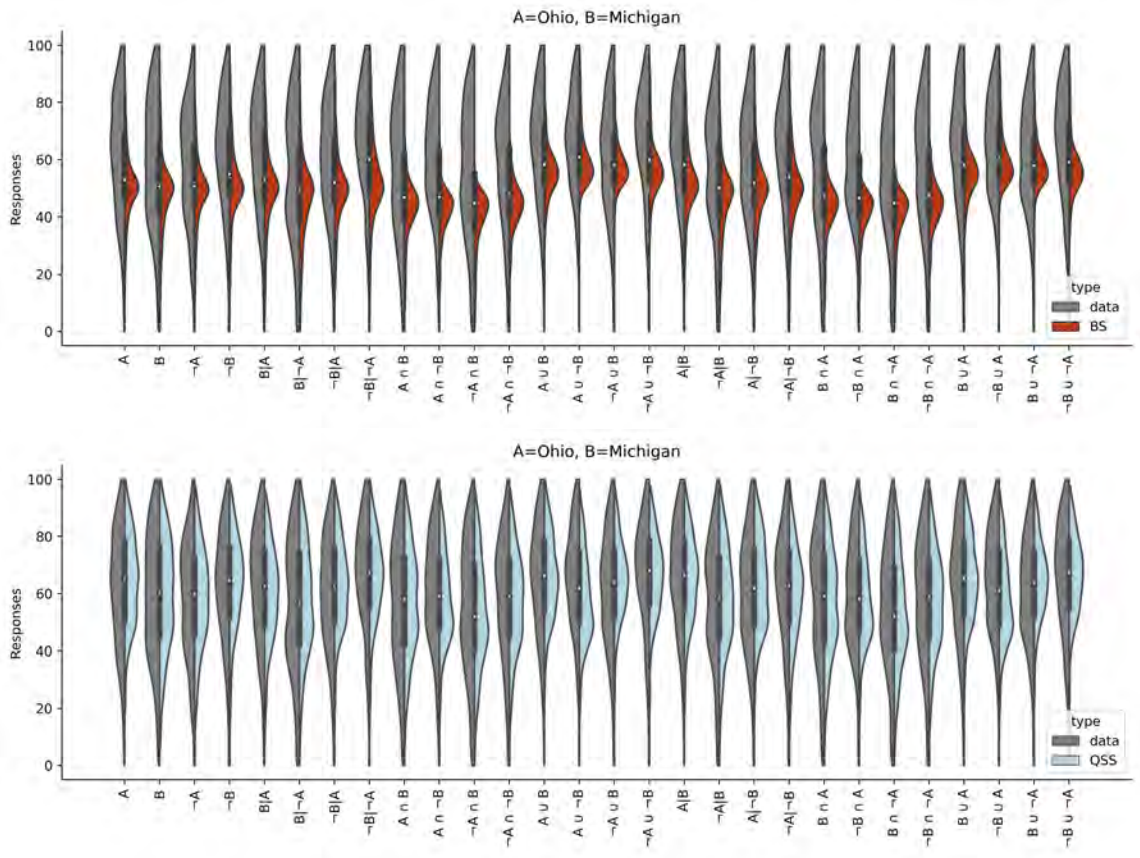


Figure 7. Violin plots showing the distribution of empirical data versus the distribution of predictions from the Quantum Sequential Sampler (top panel) and the distribution of empirical data versus the distribution of predictions of the Bayesian Sampler model (bottom panel). The black bar in the middle shows the first and third quartile of data points (25% to 75%). The data are for the $\{A, B\}$ pair of events, for the $T1$ triplet.

the one considered above (the uniform random model), but we include it here so as to follow more closely Zhu et al. (2020) and to illustrate an alternative approach to a baseline model.

Mean model predictions are shown in Figure 6, for the first triplet of events and the distribution of predictions in Figure 7, for one pair of events from the first triplet. Supplementary Material 3 (Figures S.3.1 − S.3.5) shows predictions for the second triplet and distributions for the remaining pairs. In all of these plots, only the quantum variant of the Quantum Sequential Sampler model is shown. Given how similar their BIC results are, one can expect the classical variant to have a similar mean prediction as the quantum variant. The difference between the classical variant only matters when we examine prediction at the

individual level, which we will consider more closely later. Overall, the Quantum Sequential Sampler not only makes better mean predictions, compared to both the Bayesian Sampler and the relative frequency model, but also predicts the distribution of probability judgments, across participants, reasonably better.

As an additional point, inspecting the distributions of responses allows us to consider whether there are empirical indications that the different probability terms, such as $P(A \wedge B)$ versus $P(A|B)$, were understood differently by participants. Without undertaking detailed analyses, it can be seen that there are many instances in Figure 7 where conditionals and conjunctions/ disjunctions appear to have different distributions.

Probability identities

We presented earlier the probabilistic identities derived by Costello, Watts, and Fisher (2018) and also employed by Zhu, Sanborn, and Chater (2020, Table 2). The important point is that expectation diverges, depending on whether one adopts (Bayesian) probability plus noise or quantum rules. Costello et al. (2018) originally argued that their results uniformly support their account over and above the quantum model, but their work concerned Busemeyer et al.'s (2011) model, which we pointed out is incomplete. In Figure 8, we present the results examining the Quantum Sequential Sampler, the Bayesian Sampler, and the relative frequency models in predicting these 18 identities, computed using the mean predictions across participants, as Zhu et al. (2020) did, for the first triplet (results for the second triplet are shown in Supplementary Material 3). Note, for the empirical data, we ignored order differences in conjunctions and disjunctions and simply computed the average across the two orders. For any model, since there is only a single subjective probability for conjunction and disjunction, we do not need to compute an average. As expected, the relative frequency model, which strictly follows Bayesian principles, predicts zero for all the identities, while both the Quantum Sequential Sampler and the Bayesian Sampler may predict non-zero for the identities, which is closer to what is empirically observed.

The Quantum Sequential Sampler model offers a uniformly better correspondence with the value of the identities compared to the Bayesian Sampler. One reason that the Quantum

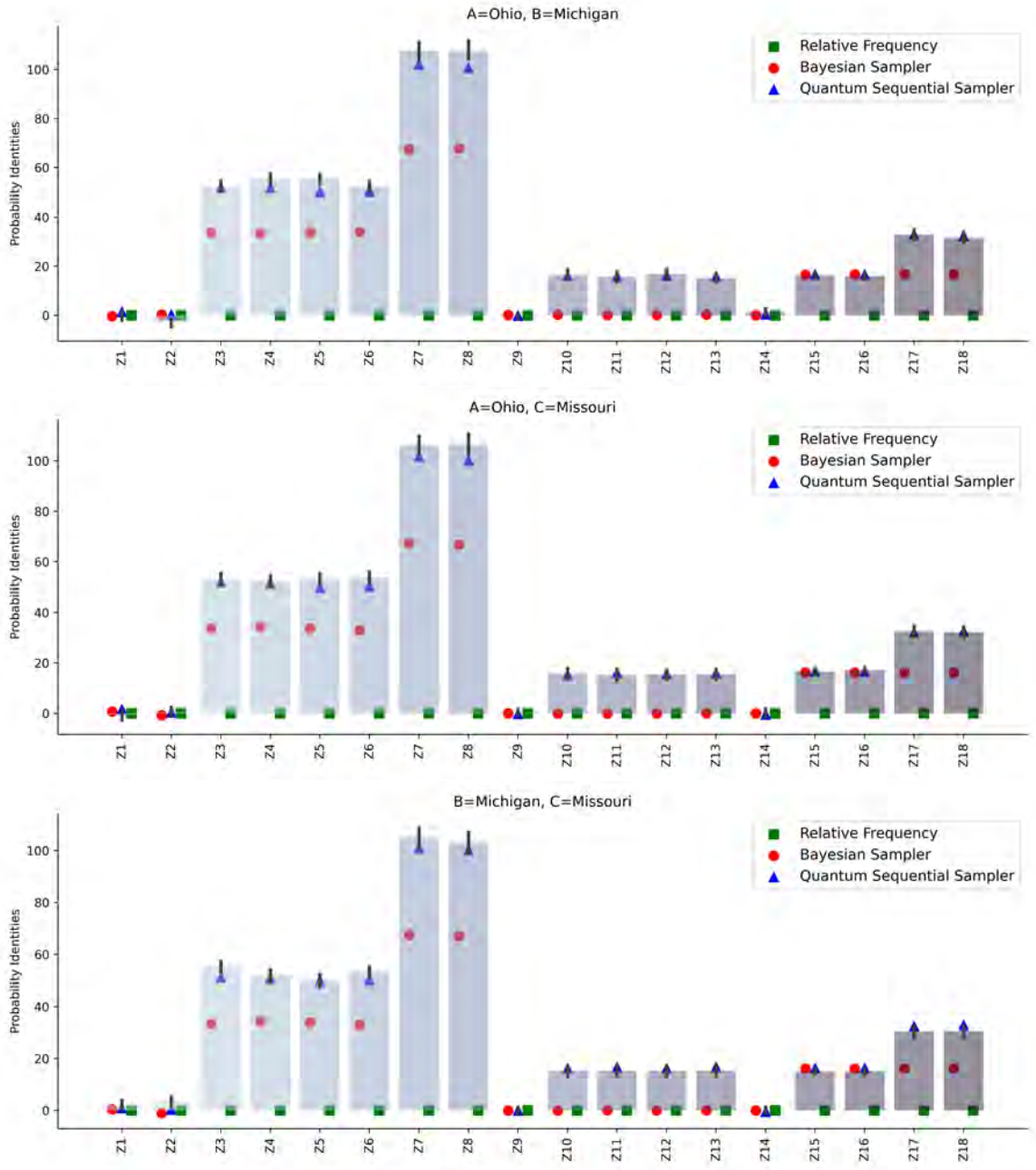


Figure 8. The empirical value of probability identities in Table 2, together with values computed from best-fit predictions of the Quantum Sequential Sampler model, the Bayesian Sampler, and the relative frequency model, averaged across all participants. The error bar in the middle shows the 95% confidence interval of the means. The results are for Triplet $T1$.

Sequential Sampler model predicts these identities better than the Bayesian Sampler is that there are some systematic differences between the observed value of these identities in our

dataset, compared with Zhu et al.'s (2020) one, as mentioned above. This potentially also explains why the Bayesian Sampler model did not perform as well in our dataset, compared to in the original Zhu et al. (2020) study. In the Quantum Sequential Sampler, the interference parameter interacts in a more complex way with the various probability terms, offering a more precise balance.

Another interesting observation about the probability identities is that the values of all these identities seem to be relatively constant, across the different question pairs we investigated. This is an unexpected result, given that there are some variations between the values of conjunctions, disjunctions, and conditionals, across the three pairs in each triplet. One reason for the constancy of the identity values might be that variations between pairs may not just be large enough in our dataset. For instance, participants might rate similarly the probabilities of the two candidates winning some states. It is an interesting question for future research whether the value of these identifies survives variation in the probability judgments, in datasets where such variation is more pronounced.

Finally, a noteworthy aspect of Costello and Watts' (2014) and Zhu et al.'s (2020) work is that they were able to analytically derive exact predictions for these identities from their models. With the Quantum Sequential Sampler, this is not possible – the model is too complex and, in the most general case, it is impossible to disentangle the influence of the two parts, quantum probabilities and sequential sampling. However, a linear approximation to the model can help somewhat in this respect, as we showed above for binary complementarity. In Appendix 3 we consider the probabilistic identities, as well as some other notable fallacies. Qualitatively, we replicate the multiplicative relations between various probability identities in Costello and Watts (2014), using a linear approximation to our model.

Binary Complementarity

As noted previously, a distinguishing feature of the Quantum Sequential Sampler is its ability to account for binary complementarity violations, a fallacy that eludes explanation by existing models. In Figure 9 and in Supplementary Material 3, we show how model predictions regarding this fallacy are in line with empirical results. We also show corresponding

predictions from the Bayesian Sampler and a simple relative frequency model. Both models are constrained to obey binary complementarity and so they fail to accurately predict the substantial violations observed in the data. As shown in Figure 9, another notable finding is the consistent overestimation effect observed across all complementary pairs. This effect is not only pervasive but also exhibits a remarkable uniformity in its magnitude: for each pair of complementary elements, the sum of their probabilities consistently approximates 1.2.

Because Zhu et al. (2020) and Costello and Watts (2014) did not find violations of binary complementarity in their studies on weather events, and given the prevalence of probability anomalies in studies concerning electoral events (e.g., Moore, 2002; Yearsley & Trueblood, 2018; Wang et al., 2014), we hypothesize that the occurrence of binary complementarity violations in our experiment may be linked to the specific characteristics of questions in election scenarios. For instance, a Trump supporter might logically assess Biden as the likely winner of Michigan, yet may be hesitant to assign a definitively low probability to Trump's victory in that state, leading to overestimation. An interesting direction for future research is to explore a wider array of question types and evaluate the incidence of binary complementarity violations at the individual level.

Comparing quantum and classical variant of Quantum Sequential Sampler

It is interesting to examine differences between the classical variant and the quantum variant of the Quantum Sequential Sampler, to pinpoint situations where quantum probability enhances the prediction of probability judgment estimates.

Before delving into the details, we consider a few preliminary points. First, theoretically speaking, quantum interference is vital for explaining the conjunction and disjunction fallacies in averaged probability judgements, akin to the role played by the additional noise term in Costello and Watts' (2014) model and the smaller sample size in Zhu et al.'s (2020) model. This can be seen in Equation 22: with constant b and α, the mean judgment $\mu_{QSS}(A \wedge B)$ exceeds $\mu_{QSS}(B)$ only when $P(A \wedge B) > P(B)$, a requirement that necessitates the use of quantum probabilities. Nevertheless, it is possible to observe conjunction and disjunction fallacies at the individual participant level, even if there are no such fallacies

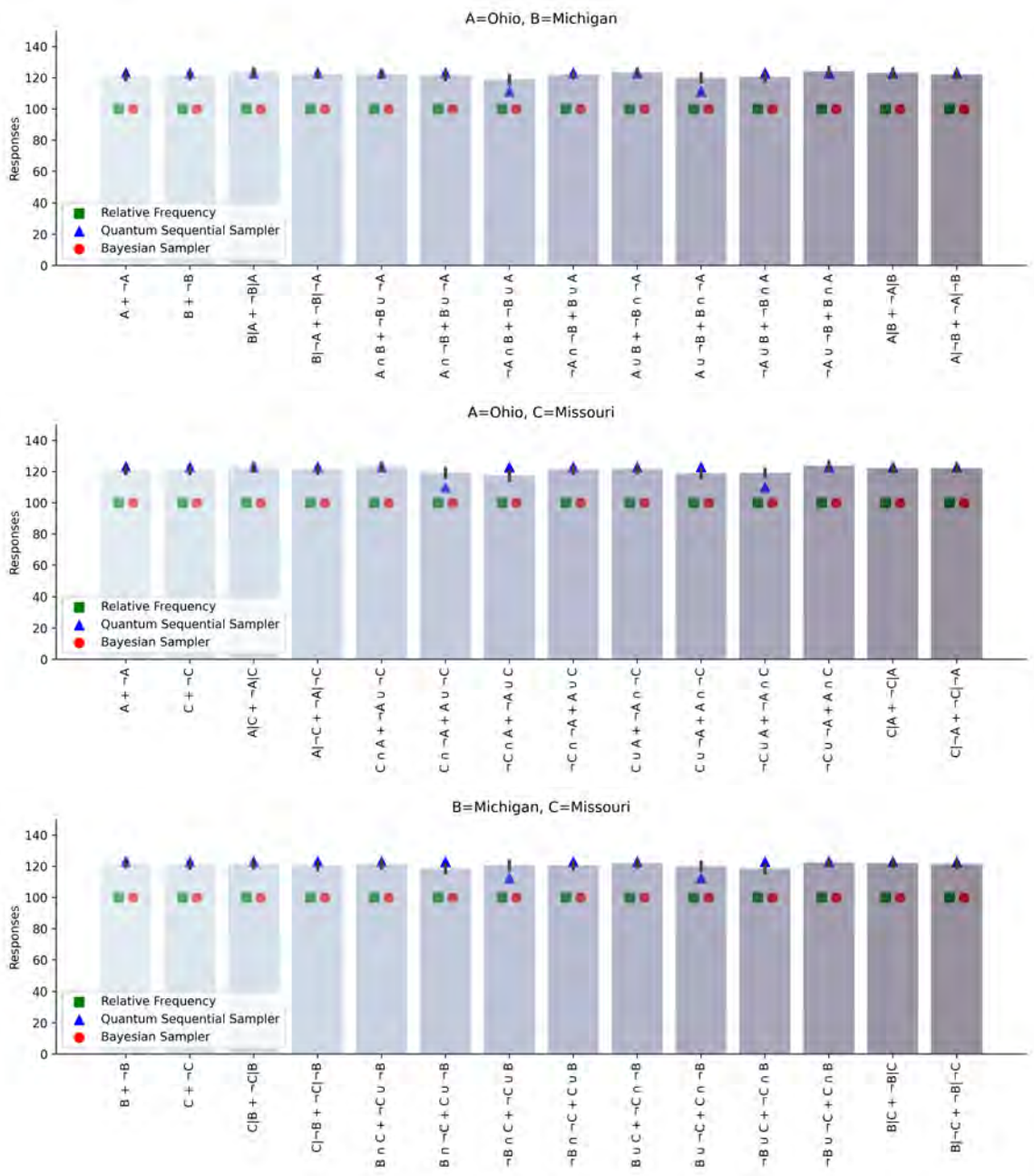


Figure 9. Empirical values for binary complementarity, together with predictions from the Quantum Sequential Sampler, the Bayesian Sampler, and the relative frequency model, averaged across all participants. The error bar in the middle shows the 95% confidence interval of the means. The results are for Triplet $T1$.

in average probability judgments (Costello & Watts, 2014). This suggests that the classical variant could by itself account for the random occurrence of conjunction fallacies as a result

of stochastic sampling processes.

Second, we mentioned earlier that for 576 out of 1162 participants, the quantum interference parameter was significant. This finding implies that an integration of sequential sampling and quantum probabilities is essential for a considerable subset of participants. However, when we analyze the averaged predictions for the entire participant pool, distinguishing between the classical and quantum variants becomes challenging (Supplementary Material 4, Figures S.4.1 − S.4.3). This difficulty is evident in the mean BIC and the outcomes of the generalization test. Although the quantum variant shows some enhancements over the classical variant (after accounting for the complexity introduced by an additional parameter), these improvements are modest in contrast to the Quantum Sequential Sampler's advantage over the Bayesian Sampler model. Additionally, while the quantum interference parameter significantly impacts half of the participants, for the remaining participants responses align well with the predictions of the classical variant.

In light of these factors, we have decided to carry out some further analyses for the 576 participants who exhibit a significant quantum interference parameter, as a way to gain insight into when the quantum interference parameter is necessary for understanding the probabilistic reasoning of these particular individuals. Below we present analyses both for the full set of 576 participants and for a smaller subset for whom the quantum advantage was established with a more strict criterion.

We first analyzed the mean predictions of the quantum and classical variants for the 576 participants for whom the significance of the quantum interference parameter was established according to the "two-sigma" criterion, that is, $p < .05$. These findings are presented in Figures S.5.1, S.5.3 in Supplementary Material 5. For this subset of participants, the quantum model demonstrates marginally superior accuracy in predicting both the mean judgments and their distribution; the degree of improvement is, however, very small.

Besides examining predictions which, according to Equation 24, represent the expected value of the final state $\phi(t)$, we also investigated the standard deviation of this final state (see Figure 10). Here, the results more clearly favor the quantum variant, which exhibits a con-

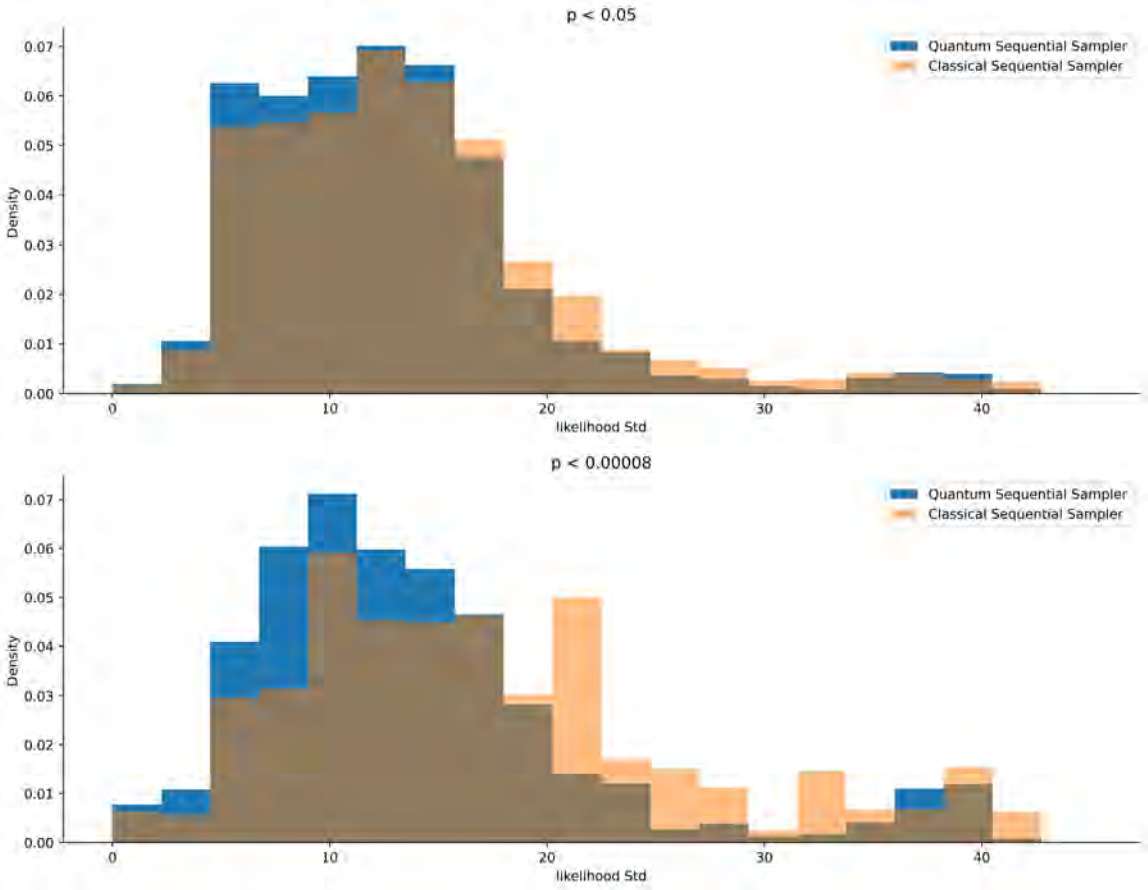


Figure 10. Histogram showing the distribution of standard deviation of the final state of the Markov process $\phi(t)$. Since the likelihood of the data given the model is directly computed from this state, the standard deviation of the state is the same the standard deviation of the likelihood distribution.

sistently smaller standard deviation for these "two-sigma" participants. To elaborate, within our framework of maximum likelihood estimation (see Equation 19), a reduced standard deviation of $\phi(t)$ implies a greater probability that the model can generate the empirical data successfully, particularly when $\phi(t)$ adheres to an approximately Gaussian distribution with prediction close to the data. Given that both model variants closely replicate the observed data and considering that $\phi(t)$ is initially constructed from a symmetric beta distribution, the quantum variant's reduced standard deviation indicates enhanced predictive capability. This extra sharpness in prediction of the quantum variant is attributed to the additional interference parameter, which allows the model to produce probabilities closer to empirical results, when these reflect Bayesian fallacies. By contrast, the classical variant must account

for such results with greater variability.

The role of the quantum interference parameter becomes more pronounced when applying the "four-sigma" criterion ($p < .00008$) to determine its significance. Under this more stringent criterion, only 95 participants exhibit a significant quantum interference parameter. For this smaller group, a clearer distinction emerges between the mean predictions of the quantum and classical variants, with the quantum variant demonstrating superior performance, as evident in Figures S.5.3, S.5.4, in Supplementary Material 5. For the "four-sigma" participants, the distribution of predictions from the classical variant are more narrow and less well aligned to empirical data. Additionally, for these participants, the quantum variant shows a more substantial improvement over the classical variant in terms of the standard deviation of $\phi(t)$ compared to the "two-sigma" participants, according to Figure 10. Clearly, in instances where predictions from the classical variant significantly deviate from probability judgments, a larger standard deviation is necessary to accommodate the empirical data.

To verify that the results from the "four-sigma" participants are not simply the product of random fluctuations within the fitting process, we replicated the fitting procedure thrice for these individuals. Each iteration yielded virtually identical G^2 values, indicating consistent outcomes. This consistency reinforces the notion that the classical variant's subpar fit for these participants is likely due to the inflexibility of the classical probability framework, rather than random errors during fitting. Note, identifying all probabilistic fallacies that challenge the classical variant is an open-ended objective. While the search for probabilistic fallacies has been intensely carried out for several decades now, it is unclear how definitive the current list is – indeed, in the present work, the newly established violations of binary complementarity played a key role.

As a further attempt to understand the behavior of "four-sigma" participants, we employed kernel density estimation to compare the predictions of classical and quantum variants against the empirical data, as illustrated in Figure 11 and detailed in Supplementary Material 6. Additionally, we present the kernel density estimation for a participant whose optimal quantum interference parameter is zero. Participants whose responses are

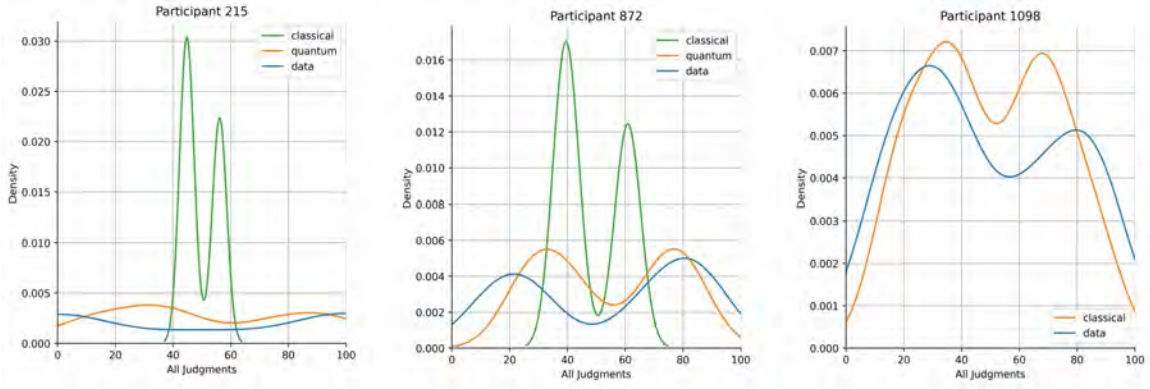


Figure 11. Kernel density estimation for observed and predicted judgments, for two quantum participants (215, 872) and one classical participant (1098), corresponding to whether the interference terms are significantly non-zero or zero in the Quantum Sequential Sampler. For the first quantum participant 215, the triplet is A = Ohio, B = Michigan, and C = Missouri, and for the the other two participants 872 and 1098, the triplet is A = Georgia, B = Montana, and C = Nevada. Note, all 78 judgments are represented in the figures.

accurately captured by the classical variant often exhibit bimodal distributions in their kernel density estimations. These two modes correspond to the complements of the two marginals, reflecting a tendency to categorize responses distinctly, as evidenced by the dual peaks in the density plots for these classical participants. Conversely, the "four-sigma" quantum participants display a broader range of responses, populating the continuum between the two peaks. This distributional characteristic could arise from systematic conjunction and disjunction fallacies, allowed by quantum interference, which can lead to predictions intermingling more closely with those of the corresponding marginals. Moreover, the kernel density plots for the quantum participants generally show higher entropy, with responses demonstrating greater variance, compared to classical participants. Visually, the density function for quantum participants appears more akin to a uniform distribution. It is important to note that these observations are qualitative and preliminary. Future research is encouraged to rigorously explore the link between the entropy and variance in the distributions of probability judgments and the quantum interference evident in participants' responses.

General discussion

We proposed a novel framework for probability judgments in probabilistic reasoning. Our aim has been to capture several insights about probabilistic reasoning, some of which have already been expressed in current formalisms, but (we think) in incomplete ways and never over a scope as encompassing as the present one. The main two insights driving our work concern the nature of subjective probabilities and the way that subjective probabilities drive observed responses.

Our first main assumption is that probabilistic reasoning reflects subjective probabilities and these subjective probabilities need to be distinguished from the observed probability ratings. This is an assumption which contrasts with heuristic or bias approaches to cognition (e.g., Hertwig et al., 2013; Kahneman et al., 1982; Nilsson et al., 2009). Without doubt, proposals for heuristics and biases have consistently captured important aspects of behavior. However, the formalization of such accounts is often limited, so that researchers have sought to incorporate such ideas into formal frameworks (whether Bayesian, as in Lieder and Griffiths, 2019, or quantum, as in Busemeyer et al., 2011).

The theoretical difference between the Quantum Sequential Sampler and extant Bayesian sampling models, notably the ones from Costello and Watts (2014) and Zhu et al. (2020) is more subtle. In all models there is some sampling process. An agent only experiences sample values and evaluates a question or rating using these values. All models assume that this sampling process produces a distribution of sample values and distribution means equal the corresponding subjective probabilities. Compared to Costello and Watts (2014) and Zhu et al. (2020), in the present work, we explored a theory for probabilistic reasoning based on a different calculus for subjective probabilities (quantum versus Bayesian) and a different sampling process (sequential sampling versus fixed sampling). As discussed, we think there is ample evidence in the literature to justify both theoretical choices.

Regarding subjective probabilities, if we accept that they have some role in probabilistic reasoning, then a question naturally arises of whether such probabilities should be Bayesian or quantum or indeed something else. There is a pervasive intuition that cognition

encompasses both an analytic/ reflective/ thoughtful mode and a heuristic/ reflexive/ spontaneous one (Kahneman, 2001). One way to develop these ideas is to assume that Bayesian versus quantum models capture an analytic versus heuristic distinction and, importantly, that the relative weight of these influences is continuous, in terms of the size of the interference parameters. One can say that our formalism proposes an infinite number of cognitive routes, from strongly Bayesian to strongly quantum, with all intermediate possibilities in between.

What is the basis for associating Bayesian with analytic and quantum with heuristic? It is uncontroversial to consider Bayesian reasoning as the absolute rational standard in reasoning, at least in most cases, e.g., excluding cases where there might be inherent contextuality (Pothos et al., 2017). Therefore, it seems that, if we could, we should just be baseline Bayesians all the time. Of course, it is well known that this is not possible, placing an extra 'cost' on any situation when we attempt to be Bayesian (e.g., Lieder and Griffiths, 2019). One way to simplify Bayesian reasoning is to use quantum probabilities. This is because quantum theory can be seen as a compartmentalized version of Bayesian reasoning, that is, Bayesian reasoning which applies only in subsets of questions (in some question space), eliminating the need for complex probability distributions (Pothos et al., 2021; Lewandowsky et al., 2002). Note, if the situation is Bayesian and we apply quantum probabilities, we might misrepresent the world – simplification usually comes at a price. In any case, we think it is reasonable to consider quantum as fulfilling the role of heuristic modes of thought and indeed there is evidence that both unfamiliarity and more reflexive modes of thinking are associated with quantum reasoning (Trueblood et al., 2017). In the present work, we observed some association between higher CRT scores (more reflective thinking; Frederick, 2005) and fewer conjunction and disjunction fallacies. We do note that these ideas are offered here as somewhat speculatively, awaiting further examination in future work.

The above argument is agnostic concerning possible influences in human probabilistic reasoning beyond either Bayesian or quantum reasoning. Indeed, it seems likely that there would be such influences, not least because it is well known that several factors can affect

the judgment process, including emotion (Bower, 1981), motivated reasoning (Kunda, 1990), values (Schwarz, 1992), and just plain biased thought (Lewandowsky et al., 2012). Knowledge about probabilities, even if biased, surely corresponds to only one influence amongst others.

Our second main assumption is that subjective probabilities are not known directly, but rather serve as drift rates guiding a sequential sampling process, corresponding to a participant trying to decide on an appropriate response. Compared to previous proposals with a sampling component (Costello & Watts, 2014; Zhu et al., 2020), the main advantage of a sequential sampling framework is that it obviates the need for an a priori commitment concerning the extent of sampling: sampling can be terminated when there is enough evidence for a particular response or more flexibly, depending on any combination of task demands which might arise after the start of the judgment process. Additionally, a response is likely to be a function of many different influences, over and above the actual subjective probabilities. While we cannot model such influences directly, we can allow for a process which distorts subjective probabilities, in some principled way, in terms of the way the parameters for the process are set up.

Model comparisons showed our proposal, the Quantum Sequential Sampler, to be superior to the Bayesian Sampler and to a classical variant of the Quantum Sequential Sampler. Even though in recent years there have been several sophisticated proposals for bounded-rational Bayesian reasoning (e.g., Dasgupta et al., 2020; Lieder & Griffiths, 2019), the Bayesian Sampler offers a full probabilistic calculus, capable for accommodating predictions for any probability question. There is a valid question concerning whether advantage of the Quantum Sequential Sampler over the Bayesian Sampler one comprehensively proves the necessity of quantum probabilities in probabilistic reasoning. We have presented several analyses which we think support this conclusion, but clearly this is an issue which cannot be definitely settled yet. For example, it is always possible that, if one of these alternative Bayesian proposals were to be more fully developed, conclusions might be different from the present ones. Overall, with increasingly complex approaches to biased probabilities, sam-

pling, and noise, there is a concern regarding the coherence of models under a particular label (Jones & Love, 2011). This is one reason why we think it is appealing to employ tools which are as standard as possible, in developing a formalism, as is the case with our use of a sequential sampling process. Sequential sampling processes have been extensively employed in cognition (e.g., Brown & Heathcote, 2008; Busemeyer & Diederich, 2009; Johnson & Busemeyer, 2005; Ratliff & Smith, 2015).

An important concern for the present results is whether the probabilistic task we employed is adequate. It seems uncontroversial that the main appeal of a formal probabilistic model (Bayesian or quantum), as opposed to one based on heuristics, is that probabilities constrain each other in a precise way. Thus, the more the probability judgments from each participant, the more precise the test for a particular model. Costello and Watts (2014) and Zhu et al. (2020) considered, for each participant, probability questions corresponding to a single pair of events – even though more than one pair were included, questions were not mixed across pairs. By contrast, we asked participants to respond to all pairwise combinations of three events, leading to 78 judgments per participant. Of course, the large number of judgments raises concerns both for the present work and previous related work (Zhu et al., 2020; Costello and Watts, 2014) that participants might fail to engage sufficiently with a task throughout its duration. Such problems have been well documented in cognitive research. In the present work, we sought a theme which, with reasonable justification, might be expected to engage participants to a greater extent than otherwise. But we have no direct measure of participant engagement, apart from a fairly soft attention check. A related, and much discussed issue, is whether participants correctly understand the algebraic meaning of the various logical connectives, e.g., conjunctions, disjunctions etc. Overall, the evidence seems to suggest that this is the case, especially when more judgments are included together (Moro, 2009).

We think exploring a larger set of inter-related judgments pays off: though this has not been our primary objective, the larger set of judgments enabled us to identify some novel empirical findings. Notably, we observed systematic overestimation effects and violations of

binary complementarity, even for marginals. There has been very sparse evidence for such violations previously (e.g., Epping & Busemeyer, 2023; Shafir, 1993) and the present results represent a main empirical contribution from this work. We also identified evidence for double conjunction fallacies (Crupi et al., 2018). These findings preclude a model based on just subjective probabilities, even quantum ones, since the quantum model of Busemeyer et al. (2011) cannot accommodate double conjunction fallacies and is limited in its capacity to accommodate violations of the law of total probability – violations of binary complementarity are not possible at all. Such conclusions bring into focus the point that investigations of probabilistic reasoning on the basis of limited test procedures are bound to offer likewise limited tests of models.

So, are we to conclude for a general recommendation of just 'more is better'? There are two potential difficulties here: first, with more elaborate question combinations, there might be genuine processing limitations, either in terms of memory or attention or even basic comprehension. For example, it is unclear how well participants might understand a question of probability conditioned on three predicates. One has to consider how often questions involving multiple predicates in complex arrangements really occur in real life. Additionally, concerning tests of probabilistic models, it is less clear how new constraints could be tested by extending the range of events beyond what has been presently employed. In this work, the point of having three pairs has been exactly that probabilistic assignment in one pair impacts on assignment in the other two pairs – because of the law of total probability and, in the Bayesian case, the requirement that all events conform to a three-way probability distribution (in the quantum case there are different constraints concerning processing order). If, for example, one were to consider four events and corresponding pairs/triplets, the tests would be of the same constraints, just over a greater range of events. It is unclear whether this matters.

We can also call into question whether a decision paradigm might be the best way to study probabilistic reasoning. Specifically, throughout this and related work, the emphasis has been on probabilistic reasoning with questions which concern the general knowledge

and experience of participants. But maybe this is a problematic approach. For example, there are certainly advantages in having participants infer probabilities directly from experimental materials (as in Fiedler et al., 2009) or if the true probabilities of the relevant events are known (as in Zhao et al., 2009). One advantage of using perceptual stimuli for probability judgments is that the same judgments can be queried repeatedly, without participants necessarily realising that this is the case (Zhao et al., 2009). Note, repeating identical probability judgments is a concern regarding the procedure of Zhu et al. (2020), as we have discussed elsewhere. Such research has led to many interesting findings, including in relation to pseudo-contingencies or illusory correlations (for the latter see Bott et al., 2021). We think there are complementary advantages between, what one might call, probability judgments on novel or meaningless stimuli and ones concerning knowledge-rich questions. In the latter case, there is less direct, experimenter control over the probabilities participants assign to different events. To use an example from the present paradigm, different participants might have wildly different notions of the probability that Biden will win in Arizona. Nevertheless, such individual differences do not impact on the formal modelling, since the question is how participant judgments for the different probability terms constrain each other. Additionally, for a task spanning several judgments (78 in the present research), employing a theme which should hold natural interest for most participants would be expected to help with engagement and attentiveness. Overall, it is reasonable that different empirical approaches might be better suited towards different empirical objectives. In the present case, the objective was to explore whether the constraints from probability theory, classical or quantum, about how different terms constrain each other are reflected in behaviour. As such, we think the choice of a current affairs theme of great topical interest, at the time of testing, is justified.

Regarding theoretical considerations, we think that the use of sequential sampling in probabilistic reasoning has considerable potential to expand this research area. Sequential sampling models have been shown to offer versatile and powerful predictions, for example, concerning task demands (such as time pressure, Ratliff and Smith, 2015), neural recording (Gold & Shadlen, 2002, 2007), in categorization (Nosofsky & Palmeri, 1997), and in per-

ceptual discrimination (Laming, 1968; Usher & McClelland, 2001). Recasting a model of probabilistic reasoning within the sequential sampling framework offers promise of extending work on fallacies in a wide scope. Additionally, a sequential sampling framework paves the way for expanding the range of dependent variables studied in probabilistic reasoning, notably response time and confidence (e.g., Pleskac & Busemeyer, 2010). Response times have not been a focus of attention in probabilistic reasoning, so this is an interesting direction for future work. This is indeed what has been partly accomplished by Zhu et al.'s (2023) work extending the Bayesian Sampler and we hope to carry out similar work for the Quantum Sequential Sampler in the future.

There are several challenges to the Quantum Sequential Sampler. First, in physics, the move to quantum theory was as difficult and counterintuitive for the scientists of the early 20th century, as it has been for psychologists about 15 years ago, when the quantum cognition program started. The adoption of quantum theory was initially driven by recognizing that in some cases the structure of the physical world conformed to the strange mathematics of quantum theory. Analogously, in psychology, the initial case for quantum cognitive models was based on the discovery that quantum interference terms sometimes provided simple and compelling explanations for the various apparent fallacies, especially in probabilistic reasoning (Pothos & Busemeyer, 2022). However, subsequently, in physics, an extensive and profound foundational debate ensued, concerning why quantum theory might sometimes offer a good description of the natural world (Hardy, 2002a). The objective has been to derive the axioms of quantum theory from some basic intuitions about nature (Hardy, 2002b). This step has been missing in psychology: if there is sometimes incompatibility in mental representations, why might this be the case? Or, put differently, could we consider what is the purpose of these quantum-like interference terms? Some researchers have attempted to develop an informational efficiency argument (Pothos et al., 2021), but much more work is needed. There are some related problems, such as the a priori determination of incompatibility, though in practice such problems can be circumvented, e.g., by empirical tests for incompatibility.

Second, the biggest challenge is to consider whether the present approach, based on a flexible mix between Bayesian and quantum influences, might itself be a special case of an even more general model. Note that, in the same way quantum theory generalizes Bayesian theory with the introduction of an interference term, it is possible to generalize quantum theory with even more powerful interference terms (Sorkin, 1994; Narens, 2014). In physics, it has not proved necessary to pursue such developments (Hardy, 2002a). Perhaps in psychology they might be more necessary? Additionally, there have been other probability frameworks, such as support theory (Tversky & Koehler, 2012). On the whole, such alternative probability frameworks have attracted less attention in the exploration of probabilistic reasoning, because they neither benefit from the normative/ formal justifications of probability frameworks proper (such as Bayesian theory) nor from the flexibility of pure heuristics and biases accounts. Nevertheless, it is possible, that there is a non-standard probability framework, which exactly captures the structure of human probability judgments, without the necessity of postulating separate influences.

Third, the adoption of quantum theory in this and other work is underwritten by the question of whether quantum theory might be needed at all. The finding that the Quantum Sequential Sampler (with interference effects) accounted for more participants than the Classical Sequential Sampler (without interference effects) indicates the necessity for using quantum probabilities. An interesting direction is to consider ways to resolve the problem of whether quantum probabilities are needed or not, even in cases for which fits with and without the quantum part are similar. More generally, research in probabilistic reasoning has shown that mimicries in more limited datasets can sometimes be resolved in extended ones. Perhaps extending the range of probabilities each participant considers might be useful , though, as discussed, it seems unclear whether it is worth pursuing more extensive datasets. An alternative direction might be manipulations on the nature of the events, e.g., whether the probability judgments concern weather events versus election events. Perhaps more extreme or incongruent events are more likely to be represented in a quantum-like manner.

Fourth, a technical consideration is that the form of the sequential sampling process need not be restricted to a Markov / diffusion model based on the Kolmogorov forward equation, as presently employed. It is possible to specify an analogous dynamical process based on quantum theory (Busemeyer et al., 2006; Kvam et al, 2015; Rosner et al., 2022). In quantum theory, instead of the Kolmogorov forward equation, one would use either the Schrödinger equation or the Lindblad equation. In such a case, a model would evolve not by sequentially sampling from a Markov process, but, instead by a quantum walk process. The quantum walk process evolves a superposition state of probability judgments across time, which offers some interesting characteristics compared to the classical approach.

Cognitive models based on quantum dynamical equations have been previously developed. Dynamical models using Schrödinger's equation display a characteristic indefinite oscillatory pattern and so they have been considered appropriate for capturing ambivalence in decision making (Kvam, Busemeyer, & Pleskac, 2021). The Lindblad equation includes a part which allows eventual stabilization of the dynamics and offers patterns more analogous to standard diffusion models (Rosner et al., 2022). However, the distinction between Schrödinger and Lindblad dynamics also involves additional assumptions concerning the nature of representations, which cannot be mapped to cognitive applications, without further theory (Asano et al., 2011). Also, in some cases, we have found that Markov dynamics captures human behavior better (Busemeyer, Wang, & Townsend, 2006). Concerning the Quantum Sequential Sampler, there was a good a priori reason why to include a part based on quantum probabilities – some of the fallacies have a fairly natural explanation employing quantum probabilities. However, a similar justification was just not available concerning the dynamical part; there was no reason to motivate the use of either the Schrödinger or the Lindblad equations, instead of the Kolmogorov forward equation (Kvam et al., 2021; Rosner et al., 2022). So, we think we are justified concerning this modeling option.

Fifth, another technical consideration is whether to employ projectors for cognitive measurements or allow for the possibility that judgments are made with POVMs. POVMs offer a mechanism for introducing noise in probabilistic calculations. In quantum models

based on just quantum probabilities, such as mechanism may be necessary, e.g., as in White et al. (2020) or Yearsley and Pothos (2016). The Quantum Sequential Sampler already incorporates a source of noise, in the form of the sequential sampling component of the model. So, could we say that the alternative source of noise, from POVMs, is not needed? Recall that the Quantum Sequential Sampler allows for both projectors and POVMs and the latter are essential depending on the value of the interference parameter in relation to its various boundary conditions. As things turned out, model fits indicated that for many participants POVMs were needed. Therefore, it seems that the two sources of noise make unique contributions to model behavior and it is not possible to subsume one into the other. A technical direction for future work is to explore whether different kinds of operators, alternative to POVMs, might allow correspondingly different conclusions.

Sixth, one could also question our use of quantum probability, because we do not test for order effects along with the conjunction and disjunction fallacies. A general point is that order effects and conjunction fallacy are explained by the same quantum mechanism. Even if we have not explored order effects in the present work, experiments have been conducted regarding the connection between conjunction fallacy and order effects in e.g. political polls that support the quantum model. For example, Trueblood and Yearsley (2017) observed significant and large order effects and conjunction fallacies in election scenarios comparing the likelihood of Trump and Hillary winning specific states. Boyer et al. (2016) presented evidence suggesting that order effects might not always manifest in the Linda problem given some specific framing of the questions. However, in other experiments of the same work, they did identify the expected order effects.

To address a potential confusion, in our experiment, we did not assess question order effects in the way of these previous studies, that is, comparing responses to questions A and B in one order versus another order. Instead, we assessed order effects in the probabilities of conjunctions and disjunctions by presenting their components in varied orders (e.g., Trump winning Ohio followed by Biden winning Michigan and vice versa). No significant difference was observed in the probability judgments for the two different orders. We postulate that this

is because, even with varied presentation orders, participants still perceived all components simultaneously and employed a consistent processing order of the two components. We suggest participants invariably process the more likely event first, a notion introduced by Busemeyer et al. (2011) for the original quantum model for probability fallacies. The lack of systematic order effects in conjunctions and disjunctions in our data offers some indirect support that this is a reasonable approach (but see Fantino et al., 1997). In any case, it is important to differentiate this from traditional order effects experiments (e.g., as in Trueblood & Yearsley, 2017, or Trueblood & Busemeyer, 2017), whereby participants respond to two separate, binary questions one at a time and question order naturally has to match processing order.

In any case, our main objective in this work was to model the judgments for all potential probability queries concerning the two candidates winning a state within the given triplets, not just conjunction and disjunction fallacies. There are many other other noteworthy results in probabilistic reasoning, like the violations of probability identities highlighted in Costello and Watts (2014), which can be tackled by our model. Note also that in our framework not all conjunction and disjunction fallacies stem from quantum probabilities. Some may arise from noisy sampling or mere chance (see Appendix 3). That is, our model's efficacy hinges on integrating both the quantum probability and sequential sampling components. The emphasis is not on validating either component individually, but on their collective ability to predict all probability judgments.

A final related point is that conjunction and disjunction fallacies would still arise from a quantum framework, even without the more likely first assumption. However, on average, such effects would be smaller. In general, we think that processing order depends on several factors, such as salience of the questions, attentional biases, or incidental processing biases specific to individual participants. With future work, we hope to elaborate both the model and the empirical paradigms, to refine our understanding of processing order effects (see Epping et al., 2022, for corresponding work relating to similarity judgments).

In conclusion, probabilistic reasoning and decision making in general have been one of

the most researched aspects of cognition – and with good reason too, given both the immense practical importance of this area and their central role in our understanding of what it means to be human. The present contribution advances this area in three directions: first, by offering a unique, precise way to incorporate Bayesian and non-Bayesian (in the form of quantum theory) influences; second, by proposing a novel process for mapping subjective probabilities to responses, based on the widely adopted sequential sampling framework; finally, by offering detailed model examinations against the largest to date dataset on human probabilistic judgments – our dataset offered new evidence for double conjunction fallacies and, importantly, violations of binary complementarity. This work offers a comprehensive response to the question of what apparent probabilistic fallacies are about: they are a combination of response biases on Bayesian probabilities (as others have noted, e.g., Costello and Watts, 2014; Dasgupta et al., 2020; Zhu et al., 2020) and quantum probabilities. We hope that further clarity concerning the nature of fallacies and the integration of models about probability judgments with sequential sampling ones will help advance judgment and decision theory in novel, exciting directions.

References

Aaronson, S. (2005). The complexity of agreement. STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing, 634-643, ACM: Baltimore MD USA.

Abelson, R. P., Leddo, J., & Gross, P. H. (1987). The strength of conjunctive explanations. Personality and Social Psychology Bulletin, 13, 141-155.

Adler, J. F. (1984). Abstraction is uncooperative. Journal for the Theory of Social Behavior, 14, 165-181.

Aerts, D., Sozzo, S., & Veloz, T. (2015). A new fundamental evidence of non-classical structure in the combination of natural concepts. Philosophical Transactions of the Royal Society A, arXiv: 1505.04981.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjuction fallacy. Cognitive Psychology, 21 (4), 515-550.

Asano, M., Ohya, M., Tanaka, Y., Basieva, I., & Khrennikov, A. (2011). Quantum-like model of brain's functioning: decision making from decoherence. Journal of Theoretical Biology, 281, 56-64.

Atmanspacher, H. & Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. Journal of Mathematical Psychology, 54, 314-321.

Aumann, R.J. (1976). Agreeing to disagree. Annual Statistics, 4, 1236-1239. Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and order of information. Medical Decision Making, 18, 412-417.

Bower, G.H. (1981). Mood and memory. American Psychologist, 36, 129-148.

Boyer-Kassem, T., Duchêne, S., & Guerci, E. (2016). Quantum-like models cannot account for the conjunction fallacy. Theory and Decision, 81, 479-510.

Brainerd, C. J., Wang, Z., Reyna, V. F., & Nakamura, K. (2015). Episodic memory does not add up: Verbatim-gist superposition predicts violations of the additive law of probability. Journal of Memory and Language, 84, 224-245.

Broekaert, J. B., Busemeyer, J. R., & Pothos, E. M. (2020). The disjunction effect in two-stage simulated gambles. An experimental study and comparison of a heuristic logistic, Markov and quantum-like model. Cognitive Psychology, 117, 101262.

Brown, S. D. & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. Cognitive Psychology, 57, 153-178.

Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. Trends in Cognitive Sciences, 19, 383-393.

Bruza, P. D., Kitto, K., Ramm, B., & Sitbon, L. (2015). A probabilistic framework for analysing the compositionality of conceptual combinations. Journal of Mathematical Psychology, 67, 26-38.

Budescu, D. V., Wallsten, T. S., & Au, W. T. (1997). On the importance of random error in the study of probability judgment. Part II: Applying the stochastic judgment model to detect systematic trends. Journal of Behavioral Decision Making, 10(3), 173-188.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. Journal of Experimental Psychology: Human Perception and Performance, 14(2), 281.

Busemeyer, J. R. & Diederich, A. (2009). Cognitive modeling. Sage Publications.

Busemeyer, J. R. & Bruza, P. (2011). Quantum models of cognition and decision making. Cambridge University Press: Cambridge, UK.

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. Journal of Mathematical Psychology, 44(1), 171-189.

Busemeyer, J. R., Wang, Z., & Townsend, J. T. (2006). Quantum dynamics of human decision-making. Journal of Mathematical Psychology, 50, 220-241.

Busemeyer, J. R., Pothos, E. & Franco, R., Trueblood, J. S. (2011) A quantum theoretical explanation for probability judgment 'errors'. Psychological Review, 118, 193-218.

Busemeyer, J. R., Wang, J., Pothos, E. M., & Trueblood, J. S. (2015). The conjunction fallacy, confirmation, and quantum theory: comment on Tentori, Crupi, & Russo (2013). Journal of Experimental Psychology: General, 144, 236-243.

Carlson, B. W., & Yates, J. F. (1989). Disjunction errors in qualitative likelihood judgment. Organizational Behavior and Human Decision Processes, 44, 368-379.

Chater N. (1996). Reconciling Simplicity and Likelihood Principles in Perceptual Organization, Psychological Review, 103, 566-591.

Costello, F. & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. Psychological Review, 121, 463-480.

Costello, F. J., & Watts, P. (2016). A test of two models of probability judgment: quantum versus noisy probability. In Proceedings of the Cognitive Science Society.

Costello, F. & Watts, P. (2018). Invariants in probabilistic reasoning. Cognitive Psychology, 100, 1-16.

Cox, R. T. (1961). The algebra of probable inference. Johns Hopkins University Press: Baltimore, MD.

Crupi, V., Elia, F., Apra, F., & Tentori, K. (2018). Double conjunction fallacies in physicians' probability judgment. Medical Decision Making, 38, 756–760.

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. Psychological Review, 127, 412-441.

de Barros, J. A. & Suppes, P. (2009). Quantum mechanics, interference, and the brain.

Journal of Mathematical Psychology, 53, 306e313.

de Finetti, B., Machi, A., & Smith, A. (1993). Theory of Probability: A Critical Introductory Treatment. New York: Wiley.

Diederich, A. (2003). MDFT account of decision making under time pressure. Psychonomic Bulletin & Review, 10, 157-166.

Dulany, D. E., & Hilton, D. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. Social Cognition, 9, 85-110.

Elqayam, S., & Evans, J. S. B. T. (2013). Rationality in the new paradigm: strict versus soft Bayesian approaches. Thinking and Reasoning, 19, 453-470.

Epping, G. P., & Busemeyer, J. R. (2023). Using diverging predictions from classical and quantum models to dissociate between categorization systems. Journal of Mathematical Psychology, 112, 102738.

Epping, G., Fisher, E. L., Zeleznikov-Johnston, A. M., Pothos, E. M., & Tsuchiya, N. (2023). A quantum geometric framework for modeling color similarity judgments. Cognitive Science, e13231.

Epping, G. P., & Busemeyer, J. R. (2023). Using diverging predictions from classical and quantum models to dissociate between categorization systems. Journal of Mathematical Psychology, 112, 102738.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. Psychological review, 101, 519-527.

Fantino, E., Kulik, J., & Stolarz-Fantino, S. (1997). The conjunction fallacy: A test of averaging hypotheses. Psychonomic Bulletin and Review, 4, 96–101.

Fernbach, P.M., & Sloman, S.A. (2009). Causal learning with local computations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 678-693.

Frederick, S. (2005). Cognitive reflection and decision making. The Journal of Economic Perspectives, 19(4), 25–42.

Frogner, C., Zhang, C., Mobahi, H., Araya, M., & Poggio, T. A. (2015). Learning with a Wasserstein loss. In C. Cartes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), Advances in neural information processing systems (pp. 2053–2061). Montreal, Canada: MIT Press.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), Subjective probability (pp. 129-161). New York: Wiley.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: models of bounded rationality. Psychological Review, 103, 650-669.

Gilboa, I. (2000). Theory of decision under uncertainty. Cambridge University Press: Cambridge, UK.

Gold, J. I. & Shadlen, M. N. (2007). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron, 36, 299-308.

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. Annual Review of Neuroscience, 30, 535-574.

Gould, S.J. (1992). Bully for brontosaurus: Further reflections in natural history. Penguin Books.

Griffiths, T.L. & Tenenbaum, J.B. (2009) Theory-based causal induction. Psychological Review 116, 661–716

Griffiths, T. L. & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. Cognitive Science, 31, 441-480.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. Topics in Cognitive Science, 7, 217-229.

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.

Hanson, R. (2006). Uncommon priors require origin disputes. Theory and Decision, 61, 319-328.

Hardy, L. (2002a). Why quantum theory?. In Non-locality and Modality (pp. 61-73). Springer, Dordrecht.

Hardy, L. (2002b). Why quantum theory? In J. Butterfield and T. Placek (Eds.) Proceedings of the NATO Advanced Research Workshop on Modality, 'robability, and Bell's theorem. IOS Press, Amsterdam; e-print quant-ph/0111068.

Haven, E. & Khrennikov, A. (2013). Quantum Social Science. Cambridge: Cambridge University Press.

Hertwig, R., Hoffrage, U., & the ABC Research Group (2013). Simple heuristics in a social world. New York: Oxford University Press.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. Psychological Bulletin, 118, 248-271.

Hilton, D. J., & Slugoski, B. (2001). Conversational processes in reasoning and explanation. Blackwell handbook of social psychology: Intraindividual processes, 181.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. Psychological Review, 116, 717.

Howson, C. & Urbach, P. (1993). Scientific reasoning: The Bayesian approach. Chicago: Open Court.

Huang, J., Busemeyer, J. R., Ebelt, Z. & Pothos, E. M. (2023). Quantum Sequential Sampler: a dynamical model for human probability reasoning and judgments. In Proceedings of the 2023 Annual Conference of the Cognitive Science Society.

Jaynes, E. T. (2003). Probability theory: principles and elementary applications v.1: The logic of science. Cambridge University Press.

Johnson, J. G. & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. Psychological Review, 112, 841-861.

Jones, M. & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. Behavioral and Brain Sciences, 34, 169, 231.

Kahneman, D. (2001). Thinking fast and slow. Penguin: London, UK.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Khrennikov, A., Basieva, I., Pothos, E. M., & Yamato, I. (2018). Quantum probability in decision making from quantum information representation of neuronal states. Scientific Reports, 8, 16225.

Kvam, P. D., Busemeyer, J. R., & Pleskac, T. J. (2021). Temporal oscillations in preference strength provide evidence for an open system model of constructed preference. Scientific Reports, 11, 1-15.

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. Proceedings of the National Academy of Sciences, 112(34), 10645-10650.

Kunda, Z. (1990). The Case for Motivated Reasoning. Psychological Bulletin 108, 480–498.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. Science, 350, 1332-1338.

Laming, D. R. J. (1968). Information theory of choice-reaction times. Academic Press.

Lebedev, A., & Khrennikov, A. (2024). Quantum-like modeling of the order effect in decision making: POVM viewpoint on the Wang-Busemeyer QQ-equality. In Infinite Dimensional Analysis, Quantum Probability and Related Topics: Proceedings of the International Conference on Infinite Dimensional Analysis, Quantum Probability and Related Topics, QP38 (pp. 123-128).

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Lee, M. D. & Wagenmakers, E. (2014). Bayesian cognitive modeling: A practical course. Cambridge University Press.

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.

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.

Lieder, F. & Griffiths, T. L. (2019). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences, 43, 1-85.

Lopez-Astorga, M., Ragni, M., & Johnson-Laird, P. N. (2021). The probability of conditionals: A review. Psychonomic Bulletin & Review, 29, 1-20.

Luttbeg, B. & 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.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. New York: Freeman.

Macdonald, R., & Gilhooly, K. (1990). More about Linda or conjunctions in contexts. The European Journal of Cognitive Psychology, 2, 57-70.

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgment. Psychological Review, 106(1), 210.

McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. Journal of Behavioral Decision Making, 15, 1-18.

McNamara, J.M., Green, R.F., & Olsson, O. (2006). Bayes' theorem and its application in animal behaviour. Oikos, 112, 243-251.

Messer, W., & Griggs, R. (1993). Another look at Linda. Bulletin of the Psychonomic Society, 31, 193-196.

Moore, D. W. (2002). Measuring new types of question order effects. Public Opinion Quarterly, 66, 80-91.

Moro, R. (2009). On the nature of the conjunction fallacy. Synthese, 171, 1-24.

Oaksford, M. & Chater, N. (2007). Bayesian rationality: The probabilistic approach to human reasoning. Oxford: Oxford University Press.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).

Narens, L. E. (2014). Probabilistic lattices: with applications to psychology (Vol. 5). World Scientific.

Nielsen, M. A., & Chuang, I. L. (2010). Quantum computation and quantum information. Cambridge university press.

Nilsson, H., Winman, A., Juslin, P., & Hansson, G. (2009). Linda is not a bearded lady: Configural weighting and adding as the cause of extension errors. Journal of Experimental Psychology: General, 138, 517-534.

Nosofsky, R. M. & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. Psychological Review, 104, 266-300.

Oaksford, M. & Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection. Psychological Review, 101, 608-631.

Paulsen, V. (2002). Completely bounded maps and operator algebras (No. 78). Cambridge University Press.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. Cognition, 120, 302-321.

Pleskac, T. J. & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychological Review, 117. 864-901.

Pothos, E. M. & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? Behavioral & Brain Sciences, 36, 255-327.

Pothos, E. M. & Busemeyer, J. M. (2022). Quantum cognition. Annual Review of Psychology, 73, 749-778.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. Psychological Review, 120, 679-696.

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.

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, 288, 20202957.

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85, 59-108

Ratcliff, R. (2006). Modeling response signal and response time data. Cognitive psychology, 53(3), 195-237.

Ratliff, R. & Smith, P. (2015). Modeling simple decisions and applications using a diffusion model. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.) The Oxford Handbook of Computational and Mathematical Psychology, pp. 35-62. Oxford: Oxford University Press.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. Trends in Cognitive Sciences, 20, 260-281.

Rosner, A., Basieva, I., Barque-Duran, A., Glöckner, A., von Helversen, B., Khrennikov, A., & Pothos, E. M. (2022). Ambivalence in cognition. Cognitive Psychology, 134, 101464.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. Trends in cognitive sciences, 20, 883-893.

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.

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.

Shafir, E. (1993). Choosing versus rejecting: why some options are both better and worse than others. Memory & Cognition, 21, 546-556.

Shafir, E. & Tversky, A. (1992). Thinking through uncertainty: nonconsequential reasoning and choice. Cognitive Psychology, 24, 449-474.

Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. Memory & Cognition, 18, 229-239.

Simon, H. A. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 69, 99-118.

Sloman, S. A. (1993). Feature-based induction. Cognitive Psychology, 25, 231-280.

Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2004). Typical versus atypical unpacking and superadditive probability judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(3), 573.

Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2004). Typical versus atypical unpacking and superadditive probability judgment. Journal of Experimental Psychology: Learning, Memory and Cognition, 30 (3), 573-582.

Sorkin, R. D. (1994). Quantum mechanics as quantum measure theory. Modern Physics Letters A, 9, 3119-3127.

Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. Trends in Cognitive Sciences, 10, 327-334.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. Cognitive Science,27, 453–489.

Tenenbaum, J.B, Kemp, C., Griffiths, T.L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. Science, 331, 1279-1285.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? Cognitive Science, 28, 467-477.

Tentori, K. (2021). What can the conjunction fallacy tell us about human reasoning? In Human-Like Machine Intelligence. Oxford University Press.

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. Journal of Experimental Psychology: General, 142, 235-255.

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.

Trueblood, J. S., & Hemmer, P. (2017). The Generalized Quantum Episodic Memory Model. Cognitive science, 41, 2089-2125.

Trueblood, J. S. & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. Cognitive Science, 35, 1518-1552.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multi-attribute linear ballistic accumulator model of context effects in multi-alternative choice. Psychological Review, 121, 179-205.

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.

Tversky, A. & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization (pp. 79-98). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjuctive fallacy in probability judgment. Psychological Review, 90, 293-315.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. Psychological Review, 101, 547–567

Tversky, A. & Koehler, D. J. (2012). Support Theory: A nonextensional representation of subjective probability. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), Heuristics and Biases: the Psychology of Intuitive Judgment, pp. 441-473. Cambridge: Cambridge Universioty Press.

Usher, M. & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychological Review, 108, 550-592.

Usher, M. & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. Psychological Review, 111, 757-769.

Valone, T. J. (2006). Are animals capable of Bayesian updating? An empirical review. Oikos 112, 252–259.

Valone, T. J. & Giraldeau, L.A. (1993). Patch estimation by group foragers: what information is used? Animal Behavior, 45, 721-728.

Von Neumann, J. (1932). Mathematical foundations of quantum mechanics.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. Management Science, 39(2), 176-190.

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. Bulletin of the Psychonomic Society, 31(2), 135-138.

Wang, Z. J., & Busemeyer, J. R. (2021). Cognitive choice modeling. MIT Press.

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. Proceedings of the National Academy of Sciences, 111, 9431-9436.

Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus and problem type. Cognition, 107, 105-136.

Westfall, P. H., Johnson, W. O., & Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. Biometrika, 84, 419-427.

White, L. C., Pothos E. M., & Jarrett, M. (2020). The cost of asking: how evaluations bias subsequent judgments. Decision, 7, 259-286.

Wojciechowski, B. W. & Pothos, E, M. (2018). Is there a conjunction fallacy in legal probabilistic making? Frontiers in Psychology, 9, article 391.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. Psychological Review, 114(2), 245-272.

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation". Organizational Behavior and Human Decision Processes, 37(2), 230-253.

Yearsley, J. M. & Pothos, E. M. (2016). Zeno's paradox in decision making. Proceedings of the Royal Society B, 283, 20160291.

Yearsley, J. M. & Busemeyer, J. R (2016). Quantum cognition and decision theories: a tutorial. Journal of Mathematical Psychology, 74, 99-116.

Yearsley, J. M. & Trueblood, J. S. (2018). A Quantum theory account of order effects and conjunction fallacies in political judgments. Psychonomic Bulletin & Review, 25, 1517-1525.

Zhu, J., Sanborn, A. N., & Chater, N. (2020). The Bayesian Sampler: generic Bayesian inference causes incoherence in human probability judgments. Psychological Review, 127, 719-746.

Zhu, J. Q., Sundh, J., Spicer, J., Chater, N., & Sanborn, A. N. (2023). The autocorrelated Bayesian sampler: A rational process for probability judgments, estimates, confidence intervals, choices, confidence judgments, and response times. Psychological review.

Appendix 1

We describe our proposal for a diffusion process, matched to the Markov one in main text. The evidence accumulation process with continuous time and continuous state space (e.g., when probability judgments are measured by real numbers) is governed by the Fokker Planck Equation, with constant drift rate δ and constant diffusion rate $D = \frac{\sigma^2}{2} > 2$ (Wang & Busemeyer ,2021):

$$\psi_t(x,t) = D\psi_{xx}(x,t) - \delta\psi_x(x,t), \tag{A.1.1}$$

where ψ is some probability density function over probability judgments. Judgments correspond to some event A, where A can be an isolated conjunct, a conjunction, disjunction, or conditional event. Equation A.1.1 can be made to depend on the subjective probability of A, $P(A)$. We are looking to solve it, by specifying δ, D, an initial condition, and two boundary conditions. We first define

$$\begin{aligned} \beta_+ &= P(A) * \alpha + c_+ \\ \beta_- &= (1 - P(A)) * \alpha + c_-, \end{aligned} \tag{A.1.2}$$

where $\alpha \geq 0$ is the drift parameter, and c_+, c_- are further defined by a free additive bias parameter b: $c_+ = \begin{cases} 1 \text{ if } b \leq 0 \\ b \text{ if } b > 0 \end{cases}$ and $c_- = \begin{cases} 1 \text{ if } b \geq 0 \\ -b \text{ if } b < 0 \end{cases}$.

Note, the definition of k ensures that β_+, β_- are always positive. Regarding an intuitive understanding of the β_+, β_- quantities, we refer readers to the description of the Markov model in main text – these parameters can be more easily understood in relation to the Markov model and, moreover, the diffusion model depends more obviously on the diffusion rate and the drift rate, which we consider next.

The diffusion rate D and the drift rate δ are defined as

$$\begin{aligned} D &= \frac{\sigma^2}{2} = \frac{\Delta^2(\beta_- + \beta_+)}{2} \\ \delta &= \Delta(\beta_+ - \beta_-). \end{aligned} \tag{A.1.3}$$

Note that $\sigma^2/2$ is just another notation for the diffusion rate (we make no further use of this quantity later on). In the above, Δ denotes the step size of the Markov process in discrete space. Since in our case the states are integers from 0 to 100, step size $\Delta = 1$.

The initial condition for ψ is assumed to correspond to a probability density function (since ψ is a probability density function too), distributed according to a symmetric beta distribution $Beta(\beta, \beta)$, with free parameter β. Note, the same distribution is also employed as the Bayesian prior for the Bayesian Sampler model (Zhu et al., 2020). When $\alpha < 1$, this initial condition corresponds to

$$\psi(x, 0) = \frac{x^{\beta-1}(1-x)^{\beta-1}}{B(\beta, \beta)},\ x \in (0, 1) \tag{A.1.4}$$

where B is the beta function. When $\alpha \geq 1$, the initial condition can be specified as

$$\psi(x, 0) = \frac{x^{\beta-1}(1-x)^{\beta-1}}{B(\beta, \beta)},\ x \in [0, 1]. \tag{A.1.5}$$

We finally state the Neumann boundary conditions for solving Equation A3.1 with a reflecting boundary condition (Bhattacharya & Waymire, 2009):

$$\begin{aligned} &\lim_{x \to 0} \psi_x(x, t) = 0 \\ &\lim_{x \to 1} \psi_x(x, t) = 0, \end{aligned} \tag{A.1.5}$$

Given all the conditions above, we can find a unique solution for the probability density function over all probability judgments ψ using numerical methods. Note that ψ is also the likelihood function of a person producing a probability judgment $d \in [0, 100]$ in the data with a response time t, that is

$$L(d, t|\text{model}) = \psi(\frac{d}{100}, t). \tag{A.1.6}$$

Since the beta distribution is not defined at 0 and 1, we need to make the approximation that $L(a, t|\text{model}) = \psi(0, t) \approx \psi(0.005, t)$ and $\psi(1, t) \approx \psi(0.995, t)$, so as to have well-behaved likelihoods at these points.

Mean and Variance

According to Bhattacharya and Waymire, (2009), the change in mean for a constant coefficient diffusion process, assuming the process exists in $(-\infty, \infty)$ and vanishes at $\pm\infty$,

is the following:

$$\begin{aligned}\frac{d}{dt}\mu(t) &= \int_{-\infty}^{\infty} x\psi_t(x,t)dx \\ &= \int_{-\infty}^{\infty} x(D\psi_{xx}(x,t) - \delta\psi_x(x,t))dx \\ &= D\int_{-\infty}^{\infty} x\psi_{xx}(x,t)dx - \delta\int_{-\infty}^{\infty} x\psi_x(x,t)dx \\ &= 0 - (-\delta) = \delta. \end{aligned} \tag{A.1.7}$$

The same holds for a discrete space Markov process, except the differential equation and integral change into a difference equation and sums. Similarly for variances:

$$\begin{aligned}\frac{d}{dt}V(t) &= \int_{-\infty}^{\infty} x^2\psi_t(x,t)dx - \frac{d}{dt}(\mu(t)^2) \\ &= \int_{-\infty}^{\infty} x^2(D\psi_{xx}(x,t) - \delta\psi_x(x,t))dx - 2\delta\mu(t) \\ &= D\int_{-\infty}^{\infty} x^2\psi_{xx}(x,t)dx - \delta\int_{-\infty}^{\infty} x^2\psi_x(x,t)dx - 2\delta\mu(t) \\ &= 2D + 2\delta\mu(t) - 2\delta\mu(t) = 2D. \end{aligned} \tag{A.1.8}$$

For a reflecting boundary, analytical solutions for the mean and variance are not always possible to derive given an arbitrary initial state. However, Equation A.1.7 and A.1.8 still makes possible a linear approximation of the behavior of the Markov process, when the process is fairly far away from the reflecting boundary.

References

Bhattacharya, R. N., & Waymire, E. C. (2009). Stochastic processes with applications. Society for Industrial and Applied Mathematics.

Appendix 2

In this section, we report the fitting results for the Zhu et al. (2020) data, considering the full Quantum Sequential Sampler (including the quantum part, that is, with non-zero interference parameters) and Quantum Sequential Sampler with only Bayesian probabilities (interference parameters set to zero).

In Zhu et al.'s procedure, participants indicated their responses by reporting actual numbers. Therefore, it is likely that responses were biased towards multiples of 5, more so than what presently observed. For this reason, and so as to make our fits more comparable to those in Zhu et al. (2020), we rearranged integers in the 0 to 100 range into categories corresponding to multiples of 5, when computing G^2, to avoid complicating model fits by this bias (which is not part of any of the models). Note, we think that using a ratings scale to assess probability judgments, as in the present case, is a more robust approach, in that the distributions of responses should be more spread out across the full range of integers.

Table A2.1 shows the fitting results. The Bayesian Sampler (slightly) outperforms the Quantum Sequential Sampler for the frosty, icy, normal, typical, and warm, snowy pairs, but not for the other two cases. Note, the models were fitted separately for each pair, because this was the approach of Zhu et al. (2020) as well. We also examine the mean predictions, distribution of predictions, and predictions of probability identities comparing the two models, with results shown in Supplementary Material 8.

We can speculate concerning the apparent advantage of the Bayesian Sampler model in this case. As mentioned in main text, we think there are three possible reasons. First, Zhu et al. utilized a repeated measures procedure to assess probability, that is, participants provided three ratings for each event. It is unclear whether multiple decisions like this should be considered identical or whether probability updating might be influencing responding, e.g., judgments sometimes change corresponding beliefs (White et al., 2020, 2014 and references therein). Repeated judgments might also introduce biases in responding. For example, in the first round, a participant might say that tomorrow will be rainy with a probability of 0.9 while in the second round this becomes 0.2. In such a case, is it that they really respond

following the beta distribution or because they are asked twice and so doubt their original answer? If a participant is asked the same question over and over again, then they might wonder whether there is something going across these identical trials (e.g., a pattern to be discovered) and act accordingly. Second, the response format in the present case was more flexible than in the work of Zhu et al. (2020). The format of typing responses into the computer in the latter case plausibly encouraged responses in multiples of 5 (when measuring probabilities in a [0,100] range), motivating an additional rounding mechanism to model responses in that work. Third, we think that the weather events in Zhu et al. (2020) are more likely to be represented in a compatible way, so that interference terms would be 0. This is because, as far we know, one way in which compatible and incompatible questions are distinguished is familiarity (Trueblood et al., 2017; Yearsley & Trueblood, 2018). Weather events are very frequently considered together, whereas there would be lower familiarity for election questions, especially concerning opposing candidates. Fourth, in Zhu et al.'s (2020) case a more limited range of judgments was employed. Perhaps the available probabilities did not provide sufficiently strong constraints, to allow a cleaner separation of the models. A final possible reason is that an election, and especially that particular election, plausibly evokes more extreme opinions, which might be inducing incompatibility. For example, it might be hard to reconcile the possibility of Biden winning Ohio with Trump winning Pennsylvania, even if both possibilities are individually reasonable – that is, in quantum-like terms, such possibilities are incompatible.

Note also that the Bayesian Sampler model, which has six parameters, was fitted to 20 data points, already doing quite well. So there is not a lot of room for the Quantum Sequential Sampler to improve fit, especially bearing in mind that it has one more parameter than the Bayesian Sampler (seven parameters in total). This underscores the importance of employing triplets of events: they offer more data from a single experiment for each individual, allowing for a more nuanced assessment of model performance. Indeed, we showed that with this more complex dataset, involving 78 judgments rather than 20 for each participant, the Bayesian Sampler model does not perform as well.

Clearly, further work is necessary before any of these possibilities is supported. At this point, we are basically unsure as to why the Bayesian Sampler shows a slight advantage over the Quantum Bayesian Sampler, for the Zhu et al. (2020) dataset, even though in the case of the new dataset we collected, the Quantum Sequential Sampler comes out ahead.

Event pair	Bayesian Sampler	Classical	Quantum
{frosty, icy}	**321.87**	327.94	326.40
{normal, typical}	**319.90**	324.32	325.87
{windy, cloudy}	321.97	**314.71**	317.94
{cold, rainy}	328.58	**322.04**	323.63
{warm, snowy}	**314.95**	324.43	320.31

Table A2.1: Fit scores (BIC) for the Bayesian Sampler model and the Quantum Sequential Sampler, with ('quantum') and without ('classical') the interference term.

References

White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: the constructive role of articulating impressions. Cognition, 133, 48-64.

Appendix 3

In this section, we examine how we could use the estimated prediction of mean probability judgment from the Quantum Sequential Sampler model, to unravel the model's mechanisms for addressing a range of probabilistic fallacies. To refresh the reader's memory, according to Equation 22, the estimated mean prediction using linear approximation is given by:

$$\mu_{QSS}(A) \approx \frac{1}{2} + 2\alpha P(A) + (b - \alpha). \tag{A.3.1}$$

Note again that despite we illustrate probabilistic fallacies using $\mu_{QSS}(A) \in [0, 1]$, the actual prediction of QSS fitted to the actual empirical data are integers from 0 to 100.

Noise Cancellation

Costello and Watts (2014) identified a pivotal result concerning noise cancellation, articulated through the probability identity:

$$Z_1 = J(A) + J(B) - J(A \wedge B) - J(B \vee A) \approx 0, \tag{A.3.3}$$

where J symbolizes the empirical mean probability judgment, intentionally distinguished in this section from the subjective probabilities P. In the subsequent analysis, we employ Equation 22 to reveal how the Quantum Sequential Sampler predicts the noise cancellation phenomenon. Given the 'more likely first' principle in conjunctions and, without loss of generality, assuming that $P(A) > P(B)$, we deduce:

$$P(A \wedge B) - P(B \vee A) = P(A \text{ and then } B) + (1 - P(\neg B \text{ and then } \neg A)). \tag{A.3.4}$$

Given that:

$$P(A) - P(A \text{ and then } B) = P(A \text{ and then } \neg B), \tag{A.3.5}$$

$$P(B) - (1 - P(\neg B \text{ and then } \neg A)) = 1 - P(\neg B) - 1 + P(\neg B \text{ and then } \neg A)$$
$$= -P(\neg B \text{ and then } A), \tag{A.3.5}$$

it is evident that:

$$P(A)+P(B)-P(A\wedge B)-P(B\vee A) = P(A \text{ and then } \neg B)-P(\neg B \text{ and then } A) = o, \tag{A.3.7}$$

where o is the quantum interference parameter. Consequently:

$$\begin{aligned}
&Z_1 = \mu_{QSS}(A) + \mu_{QSS}(B) - \mu_{QSS}(A \wedge B) - \mu_{QSS}(B \vee A) \\
&\approx \frac{1}{2} + 2\alpha P(A) + (b - \alpha) + \frac{1}{2} + 2\alpha P(B) + (b - \alpha) - \\
&\left(\frac{1}{2} + 2\alpha P(A \wedge B) + (b - \alpha)\right) - \left(\frac{1}{2} + 2\alpha P(B \vee A) + (b - \alpha)\right) \\
&= 2\alpha(P(A) + P(B) - P(A \wedge B) - P(B \vee A)) \\
&= 2\alpha o \approx 0 \text{ (when } o \approx 0 \text{ or } 2\alpha \ll \left|\frac{1}{o}\right|). \qquad \text{(A.3.8)}
\end{aligned}$$

When o equals zero, the system aligns with the classical (Bayesian) framework, rendering $J(A) + J(B) - J(A \wedge B) - J(B \vee A) \approx 0$. This is consistent with predictions made by the probability plus noise model when $\Delta d \approx 0$.

It is important to acknowledge that for substantial values of α (keeping in mind that $\alpha \geq 0$), the model is likely close to the reflecting boundary, where the behavior of the means becomes hard to predict. Therefore, it is not valid to assert that noise increases monotonically as a function of α or, most pertinently, as a function of time t, which is incorporated into α.

Probability Identity Violation

Another important result for the Bayesian Sampler model is the identity:

$$Z_5 = J(A \wedge B) + J(A \wedge \neg B) - J(A) \neq 0, \qquad \text{(A.3.9)}$$
$$Z_7 = J(A \wedge B) + J(A \wedge \neg B) + P(B \wedge \neg A) - J(A \cup B) \neq 0, \qquad \text{(A.3.9)}$$

where

$$Z_7 \approx 2Z_5. \qquad \text{(A.3.10)}$$

For the Quantum Sequential Sampler suppose without loss of generality that $P(A) > P(B)$. In the case of $P(A) > P(\neg B)$:

$$
\begin{aligned}
Z_5 &= \mu_{QSS}(A \wedge B) + \mu_{QSS}(A \wedge \neg B) - \mu_{QSS}(A) \\
&\approx \frac{1}{2} + 2\alpha(P(A \text{ and then } B) + P(A \text{ and then } \neg B) - P(A)) + (b - \alpha) \\
&= \frac{1}{2} + (b - \alpha), \qquad \text{(A.3.11)} \\
Z_7 &= \mu_{QSS}(A \wedge B) + \mu_{QSS}(A \wedge \neg B) + \mu_{QSS}(B \wedge \neg A) - \mu_{QSS}(A \cup B) \\
&\approx 1 + 2\alpha(P(A \text{ and then } B) + P(A \text{ and then } \neg B) \\
&- P(B \text{ and then } \neg A) - 1 + P(\neg B \text{ and then } \neg A)) + 2(b - \alpha) \\
&= 1 + 2\alpha(P(A \text{ and then } B) + P(A \text{ and then } \neg B) - P(A \text{ and then } \neg B) - o \\
&+ P(\neg A \text{ and then } \neg B) + o - 1)) + 2(b - \alpha) \\
&= 1 + 2(b - \alpha). \qquad \text{(A.3.11)}
\end{aligned}
$$

And thus when $P(A) > P(\neg B)$:

$$Z_7 \approx 2 * Z_5, \qquad \text{(A.3.12)}$$

Similarly, when $P(A) < P(\neg B)$,

$$
\begin{aligned}
Z_5 &\approx \frac{1}{2} + (b - \alpha) + 2\alpha o, \qquad \text{(A.3.13)} \\
Z_7 &\approx 1 + 2 * (b - \alpha), \qquad \text{(A.3.13)}
\end{aligned}
$$

so

$$Z_7 \approx 2 * Z_5 \text{ (when } o \approx 0 \text{ or } 2\alpha << |\frac{1}{o}|). \qquad \text{(A.3.14)}$$

Similarly, one can show that

$$
\begin{aligned}
Z_8 &\approx 1 + 2(b - \alpha) + 2\alpha o \approx 2 * Z_5 \text{ (when } o \approx 0 \text{ or } 2\alpha << |\frac{1}{o}|), \qquad \text{(A.3.15)} \\
Z_6 &\approx Z_5 - 2\alpha o. \qquad \text{(A.3.15)}
\end{aligned}
$$

Conjunction and Disjunction Fallacy

The study of conjunction and disjunction fallacies is crucial in understanding probabilistic reasoning errors. These fallacies can manifest in various ways, including random occurrences in individual participants' responses and as systematic biases in the mean probability judgments. The Bayesian Sampler model addresses these fallacies by introducing an

additional parameter, N', representing a reduced sample size for evaluating conjunctions and disjunctions. On the other hand, the probability plus noise model accounts for these fallacies by incorporating an error propagation parameter, Δd, corresponding to higher error for conjunctions and disjunctions. In the case of the Quantum Sequential Sampler, the phenomena of conjunction and disjunction fallacies are elucidated through the quantum interference parameter o. Consider a scenario where $P(A) > P(B)$:

$$\begin{aligned}\mu_{QSS}(B) - \mu_{QSS}(A \wedge B) &\approx 2\alpha\left(P(B) - P(A \text{ and then } B)\right) \\ &= 2\alpha\left(P(B) - P(B \text{ and then } A) - o\right) \\ &= 2\alpha\left(P(B \text{ and then } \neg A) - o\right). \end{aligned} \tag{A.3.15}$$

In cases where $o < P(B \text{ and then } \neg A)$, for instance when $o = 0$ and $P(B \text{ and then } \neg A) > 0$, we observe that $\mu_{QSS}(B) - \mu_{QSS}(A \wedge B) > 0$. However, if $o > P(B \text{ and then } \neg A)$, this sets the stage for a potential conjunction fallacy. Analogous reasoning applies to disjunctions:

$$\mu_{QSS}(A \vee B) - \mu_{QSS}(B) \approx 2\alpha\left(P(\neg B \text{ and then } A) - o\right), \tag{A.3.16}$$

where a disjunction fallacy can occur if $o > P(\neg B \text{ and then } A)$.

When will Quantum Sequential Sampler be completely normative?

Much like the Bayesian Sampler Model and the probability plus noise model, the Quantum Sequential Sampler is capable of predicting participants' probability judgments to be completely consistent with classical Kolmogorov axioms. Specifically, this occurs when $\alpha = \frac{1}{2}$ and both o and b are set to zero. For any given event A, this situation can be illustrated as follows:

$$\mu_{QSS}(A) \approx \frac{1}{2} + 2 \cdot \frac{1}{2} \cdot P(A) - \frac{1}{2} = P(A). \tag{A.3.17}$$

The same hold true for conjunctions, as when $o = 0$

$$P(A \wedge B) = P(A \text{ and then } B) = P(B \text{ and then } A), \tag{A.3.18}$$

and vice versa for disjunctions.

Quantum Sequential Sampler: Supplementary Materials

Jiaqi Huang* and Jerome R. Busemeyer*
Indiana University
Zo Ebelt and Emmanuel M. Pothos
City, University of London

Supplementary Material 1

In this section, we describe the two pilot experiments which were conducted to identify judgments and materials more likely to challenge classical constraints.

Pilot Experiments 1, 2

Participants. For Pilot Experiment 1, 98 participants (70 male) from the USA were recruited on Amazon Mechanical Turk, to take part in a linked Qualtrics experiment. No other restrictions apart from location and the mandatory minimum age of 16 years were placed on participation. Each individual was paid $1.75 to take part in the experiment, which was approximately 20 minutes long. To identify participants disengaged with the task, an attention check question was used (the same one as we used for the experimental investigation, described in the main text). Twenty-two participants failed to answer the attention check correctly and were excluded from any analyses. Another two participants were excluded, despite giving correct answers to the attention check question, since they failed to complete the task in a meaningful way by either rating all probabilities 100% or 0%. Hence the final sample size was reduced to 74 participants (52 male). 73 out of 74 participants were at least 25 years of age and therefore eligible to vote in the USA.

For Pilot Experiment 2, 56 participants (25 male) from the USA were recruited in the same manner as above using Amazon Mechanical Turk. This was a shorter experiment,

lasting approximately 10 minutes, for which participants were paid $ 1.00 each. Again, participants failing a simple attention check question were excluded from subsequent analyses. Two participants were excluded in this way, resulting in a final sample of 54 (24 male). 49 out of those participants were at least 25 years of age and as such eligible to vote in the USA.

Method. For Pilot Experiment 1, participants were asked to provide 68 probability judgments each on the likelihood of one or both USA presidential candidates (Trump and Biden) winning the popular vote in certain states in the USA. Since the purpose of this experiment was to identify triplets of states that make conjunction fallacies more likely, four triplets of states were tested against each other:

- $T1$: Ohio, Missouri, Michigan
- $T2$: North Carolina, South Carolina, Pennsylvania
- $T3$: Georgia, Montana, Nevada
- $T4$: Ohio, North Carolina, Florida

Note, the reason why we examined triplets of states was because we intended the main manipulation to involve three events and their negations. The three basic events were Biden to win in state A, B or C, with the negation of these three events corresponding to Trump winning in these states. Three events would allow better tests of Bayesian constraints, for example, all two-way conjunctions in Bayesian theory are constrained by three-way conjunctions.

Participants were asked to judge the likelihood of possible combinations of events just for conjunctions, thus there were 12 conjunctions for each triplet, as well as 20 marginals, one for each state and candidate. Marginals and conjunctions were answered in two separate blocks and all questions were randomized within blocks. Examples for the questions used can be found in Figure S.1.1. All ratings were given by adjusting a slider between 0% and 100% in 1% increments, allowing for a total of 101 possible ratings along the scale. In the question text presented to participants, the candidate's name was highlighted in the color usually associated with the republican (red) and democratic (blue) parties. This choice was

made to enhance readability and allow participants to perceive the relevant information in the questions correctly, since the repetitive nature of the questions might cause attention to decline. Analogously, the 'and' in conjunctions and the 'or' in disjunctions were presented in boldface, so as to make them more noticeable to participants. All participants also answered three more questions corresponding to the Cognitive Reflection Test (CRT; Frederick, 2005), which were used to assess cognitive style. The version of the CRT used in this experiment included the following questions: (a) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (b) If it takes five minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets? (c) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Pilot Experiment 2 was nearly identical to Pilot Experiment 1 and aimed to test an additional triplet, triplet 5, which was:

- $T5$: Ohio, Wisconsin, Florida

Result

The objective of the analyses was to establish which triplets of states were more likely to lead to conjunction fallacies. Straightforwardly, a conjunction fallacy is detected, if the conjunction of two events is rated more likely than one of the marginals. This means that a conjunction fallacy is found if at least one marginal satisfies this condition. Clearly, since in these pilots we are not concerned with double conjunction fallacies, it suffices to compare the rating of a conjunction against the marginal with the lower rating. Accordingly, one-sided, paired Bayesian t-tests were conducted on all possible pairs of conjunctions and their corresponding lowest marginals. These t-tests are not correct inferential tests for the presence or absence of conjunction fallacies, because using the 'lower than either marginal' function can lead to a biased estimate of conjunction fallacy rates, if probabilities are noisy. However, here, we are only interested in whether there appear to be more conjunction fallacies for some candidate triplets vs. others, so the t-tests are useful (in main text, we

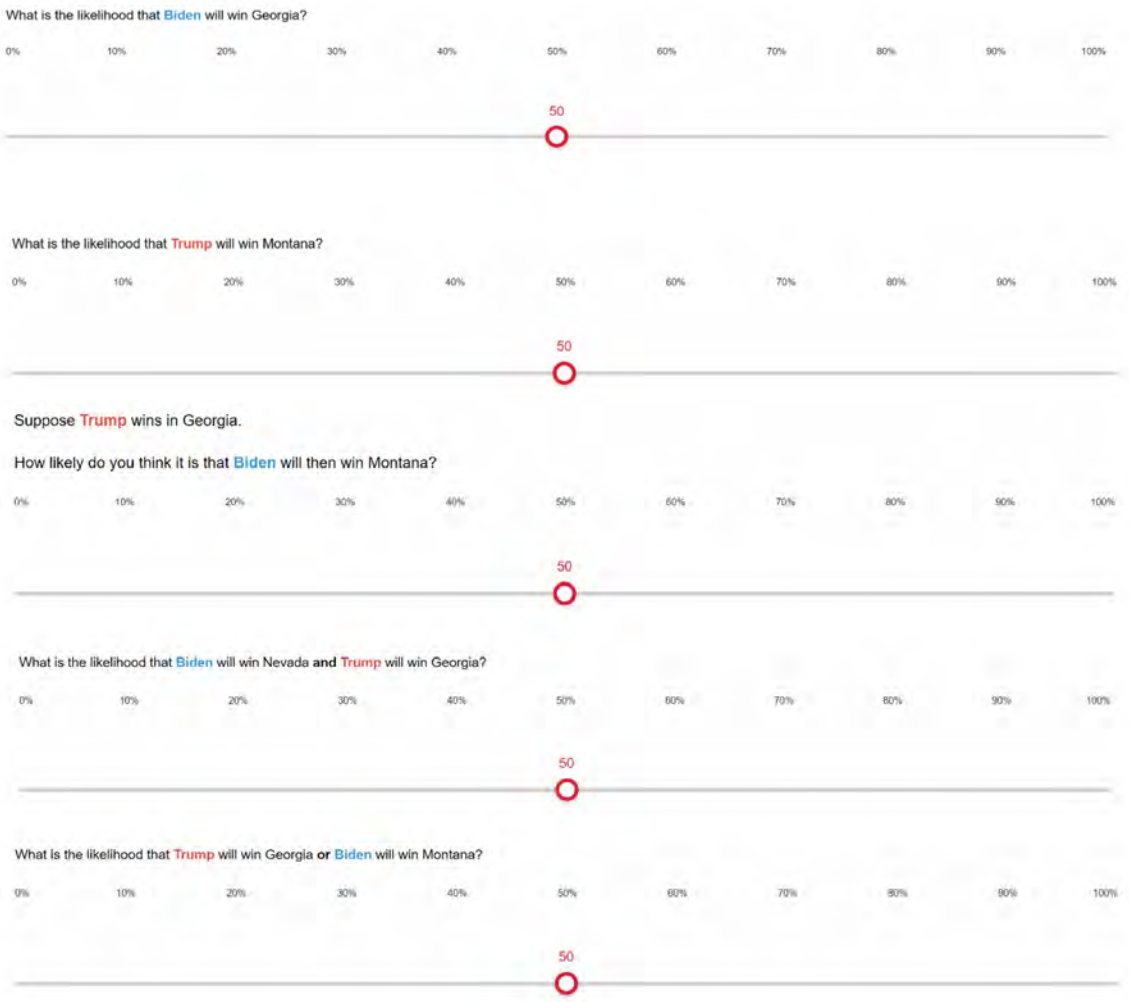


Figure S.1.1. Example questions for marginal, conditional, conjunction and disjunction probabilities.

rely directly on model fits).

We first consider the results of Pilot Experiment 1. Some conjunction fallacies were detected for all four triplets, as seen in Figure S.1.3. The results are mixed. For all triplets, we observed few significant results (i.e., conjunction fallacies), for questions that only consider one candidate at a time (these questions appear in the first block, for each triplet, in Figure S.1.3). A striking exception concerns the highly significant conjunction fallacy for the question "Trump wins Ohio & Missouri" (triplet 1, $BF_{10} = 460$). By contrast, when both candidates were mentioned in a conjunction (see the second block of questions, for

each triplet, in Figure S.1.3), there was a much higher rate of conjunction fallacies. The greatest number of conjunction fallacies was observed for triplet 4. However, for that triplet, the evidence for conjunction fallacies in cases when only one candidate was considered at a time was very weak (Bayes factors ranging between 1.50 and 6.02). Triplets 2 and 3 fall in between the extremes of triplet 1 and triplet 4, with three significant conjunction fallacies respectively and Bayes factors ranging between 2.66 and 135 for triplet 2 and between 1.49 and 8.04 for triplet 3. Note that all the figures here also contain the results for triplet 5, for ease of comparison, which will be discussed in the next section.

It is possible that the preponderance of conjunction fallacies is impacted upon by style of thinking, more reflective vs. less reflective, and previous research has offered some corresponding evidence (Yearsley & Trueblood, 2018). To examine this possibility, we computed the CRT score from each participant, which was a number between 0 and 3, depending on the questions answered correctly by each participant. In our sample, 21 participants scored 0 points, 16 participants scored 1 point, 24 scored 2 points and the remaining 13 scored 3 points. We also computed an approximate measure of each participant's tendency to produce conjunction fallacies, using the equation (as in Yearsley & Trueblood, 2018), noting again that this equation is problematic, if probability estimates are noisy (it is not used in the main part of our investigation; it is only used here, in the context of these pilots):

$$CF_{measure} = \frac{1}{\sigma} \sum_{i,j} max(P(\text{state } i \cap \text{state } j) - min(P(\text{state } i), P(\text{state } j)), 0) \qquad (1)$$

This equation offers a heuristic measure of the tendency for a participant to produce conjunction fallacies, since it increases with every instance of a conjunction fallacy, while it stays constant (0 is contributed to the sum), when there is no conjunction fallacy. The measure is normalized by σ, the standard deviation of the probability judgments of each participant, to take into account participant variability in the utilization of the ratings scale.

A Bayesian regression analysis was conducted between the $CF_{measure}$ and the CRT, yielding positive evidence for a correlation between the two variables ($r = -.35, BF_{10} = 14.1$), such that higher CRT scores were associated with larger values in the $CF_{measure}$. Post-

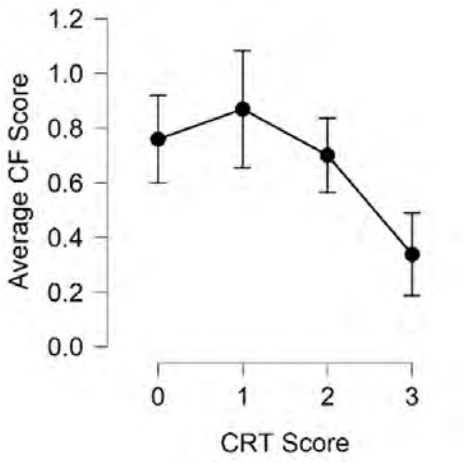


Figure S.1.2. This figure shows the $CF_{measure}$ against different levels of the CRT score. It can be seen that participants with a CRT score of 3 were associated with a weaker $CF_{measure}$, compared to participants in other groups.

hoc Bayesian t-tests were conducted to unveil the exact relationship between conjunction fallacies and CRT scores. We observed strong evidence that participants with a CRT score of 3 showed weaker conjunction fallacies compared to any other group ($BF_{10} = 17.8, BF_{10} = 34.7$ and $BF_{10} = 10.9$, respectively), see Figure S.1.2. This finding is consistent with expectation from Yearsley and Trueblood (2018), since a higher CRT score would, generally, be indicative of more reflective thinking.

Given these results, it is possible that a cleaner picture regarding the preponderance of conjunction fallacies for different triplets might be obtained if we excluded participants with a CRT score of 3, since these participants are more likely to avoid apparent classical fallacies. Repeating the above analysis with this exclusion restricted the sample to 61 participants. As shown in Figure S.1.4 (new results are indicated by shading), for all triplets there was an additional significant conjunction fallacy, with weak to strong evidence (Bayes factors ranging between 1.90 to 9.03). For triplet 3, we observed two further conjunction fallacies, but one of the previously noted ones dropped below the significance threshold.

The strength of different conjunction fallacies, which can be heuristically approximated using the $CF_{measure}$, is of particular interest for testing formal probabilistic models (both Bayesian and quantum, since, even though conjunction fallacies are allowed in the latter

case, they are bounded and so particularly strong conjunction fallacies might be outside quantum models, Yearsley & Trueblood, 2018).

The results from Pilot Experiment 1 offered a somewhat unclear picture regarding the triplets of states more likely to produce strong conjunction fallacies. Triplets 1 and 3 produced the highest (average) $CF_{measure}$, while triplets 2 and 4 offered the strongest results, in terms of the significance of Bayesian tests comparing probabilities for conjunctions against marginals.

For Pilot Experiment 2, as can be seen in Figures S.1.3, S.1.4, triplet 5 was associated with fewer significant instances of conjunction fallacies and weaker $CF_{measure}$. In this case, excluding participants with CRT = 3 decreased the number of significant conjunction fallacies, contrary to what we observed for triplets 1-4.

Taking together all results, a few options present themselves for reasonable expectation regarding the emergence of results problematic for a classical perspective. Bearing in mind that the sampling for the main experiment would be much larger than for the pilot studies, (somewhat arbitrarily) we placed more faith on the higher $CF_{measure}$ values observed for triplets 1 and 3 and chose these for the next steps:

- $T1$: Ohio, Missouri, Michigan
- $T3$: Georgia, Montana, Nevada

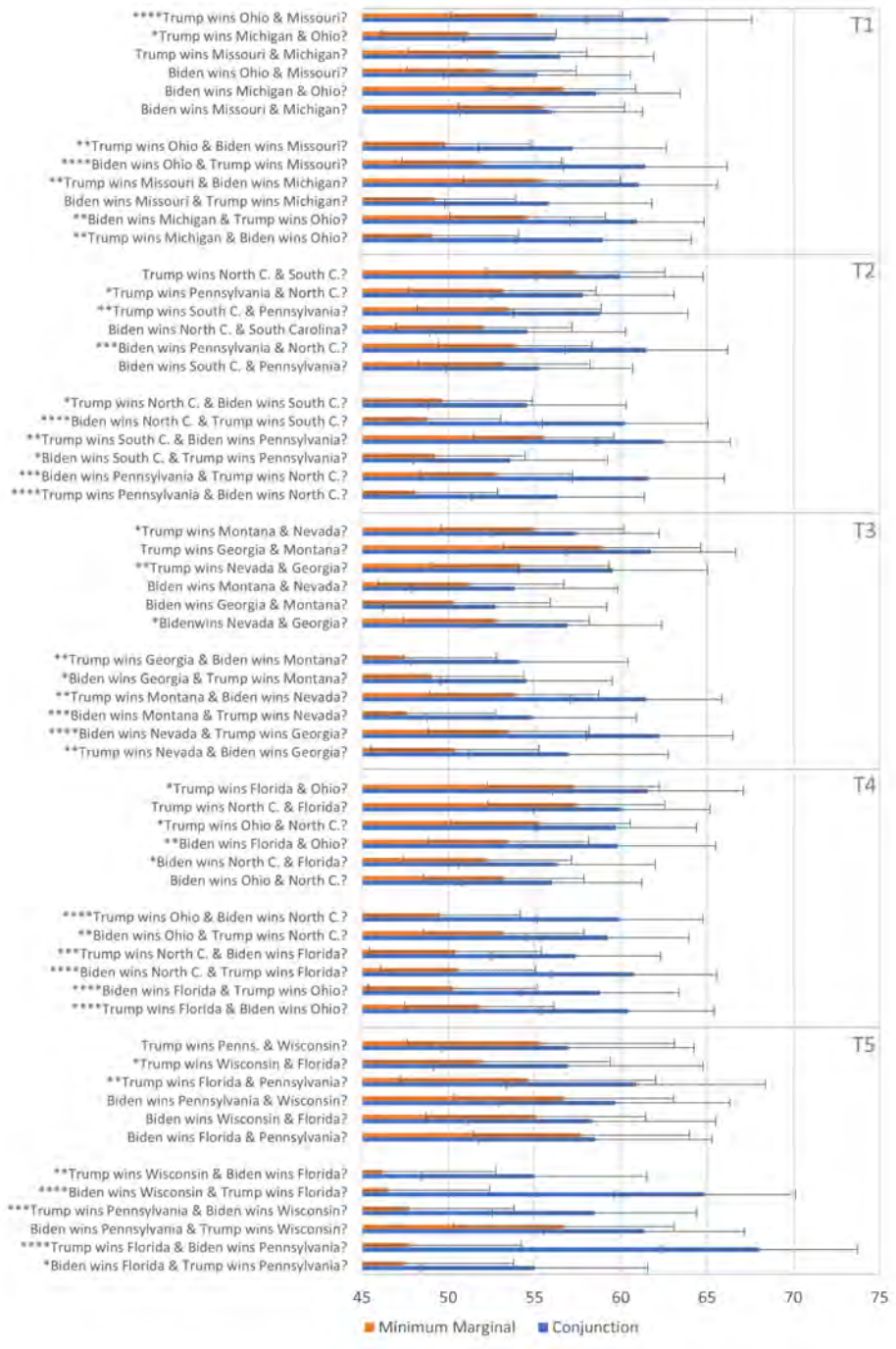


Figure S.1.3. Average conjunction probabilities (blue lines) compared to the lower corresponding marginal (red lines) for the four triplets in Pilot Experiment 1 (triplets 1-4) and Pilot Experiment 2 (triplet 5). We indicate with stars the BF_{10} evidence strength (* = weak evidence, ** = positive evidence, *** = strong evidence, **** = very strong evidence).

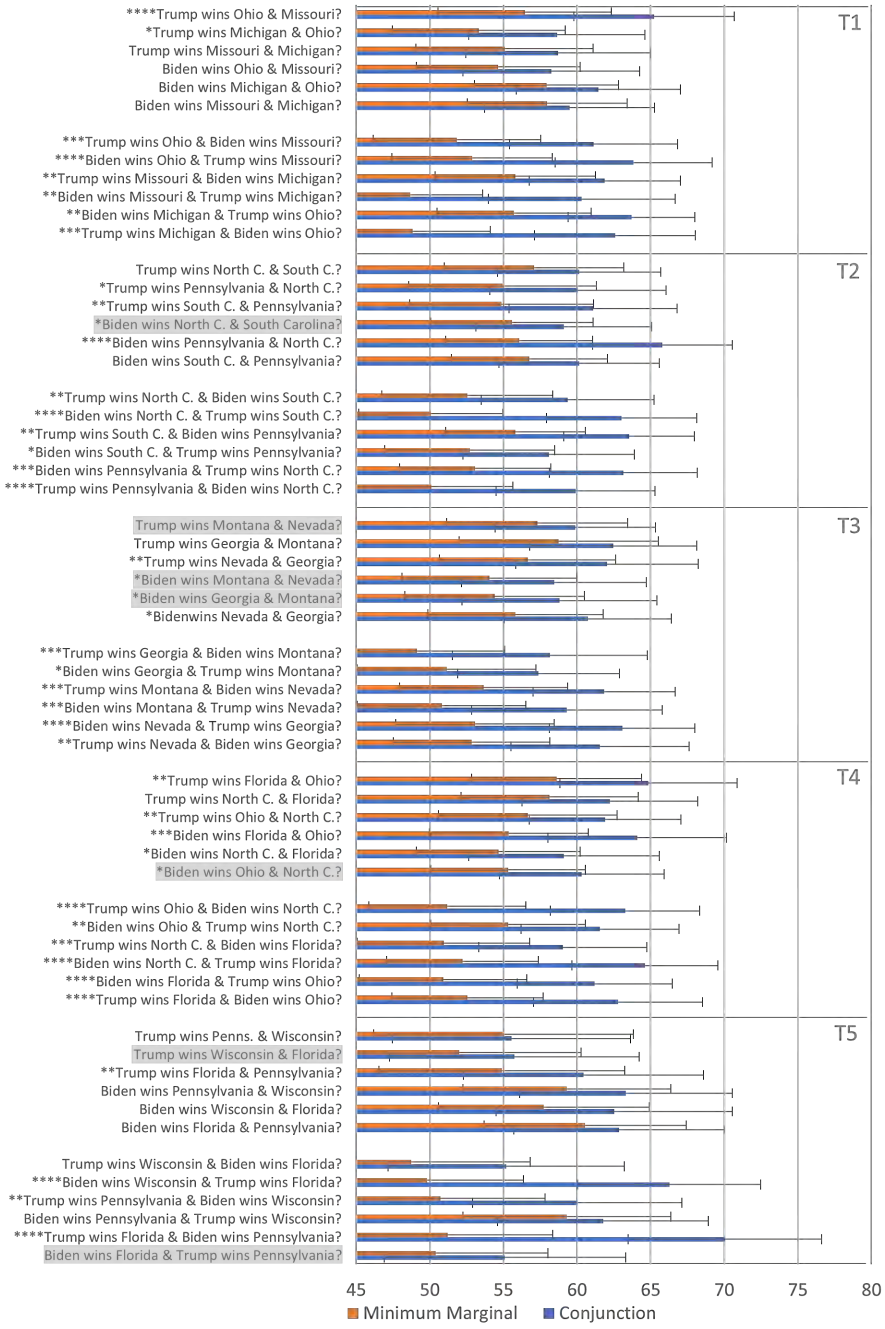


Figure S.1.4. Average conjunction probabilities (blue lines) compared to the lower corresponding marginal (red lines) for the four triplets in Pilot Experiment 1 (triplets 1-4) and Pilot Experiment 2 (triplet 5), when excluding participants with CRT = 3. We indicate with stars the BF10 evidence strength (* = weak evidence, ** = positive evidence, *** = strong evidence, **** = very strong evidence). Shading indicates differences relative to Figure S.1.3.

Supplementary Material 2

In this section, we offer some additional notes, figures, and tables, related to the standard statistical analyses of the behavioral results in main text. Note that in some cases it is less (or not at all) relevant to separate out results by triplet and so, where sensible, for brevity, we present combined results.

		Prior Odds	Posterior Odds	$BF_{10,U}$	error%
0	1	0.414	0.527	1.273	2.583×10^{-5}
1	2	0.414	652643	1.576×10^{6}	8.045×10^{-13}
2	3	0.414	1787	4314	1.384×10^{-8}

Table S2.1: Linear repeated contrasts and Bayesian post-hoc tests for the effect of CRT scores on the frequency of conjunction fallacies. Note: the posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy $(0, r = \frac{1}{\sqrt{2}})$ prior. The "U" in the Bayes factor denotes that it is uncorrected.

		Prior Odds	Posterior Odds	$BF_{10,U}$	error%
0	1	0.414	0.043	0.104	3.68×10^{-4}
1	2	0.414	0.261	0.631	9.123×10^{-6}
2	3	0.414	2.391×10^{6}	5.772×10^{6}	8.616×10^{-12}

Table S2.2: Linear repeated contrasts and Bayesian post-hoc tests for the effect of CRT scores on the frequency of disjunction fallacies. Note: the posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy $(0, r = \frac{1}{\sqrt{2}})$ prior. The "U" in the Bayes factor denotes that it is uncorrected.

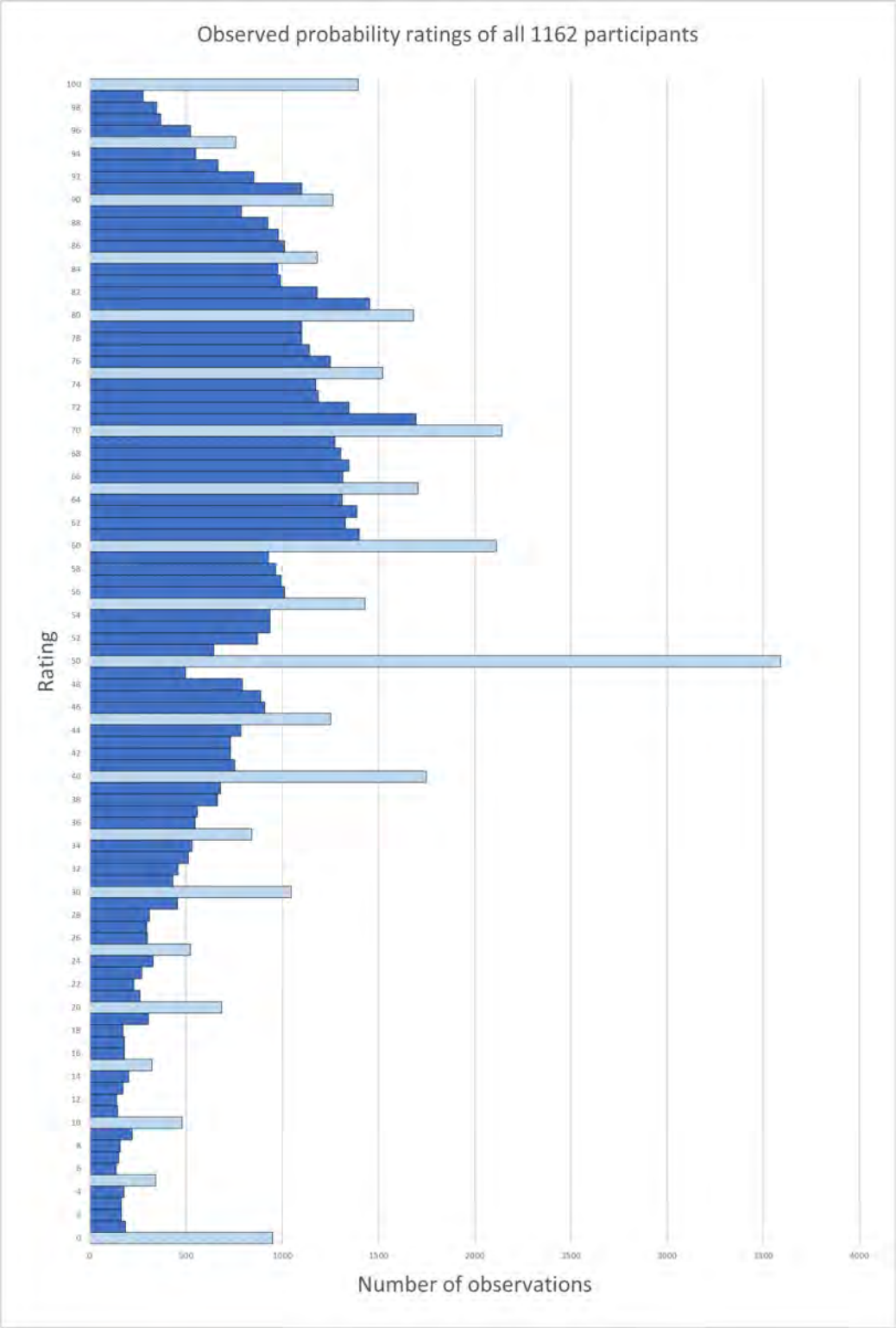


Figure S.2.1. A histogram of all observed probability ratings in the experiment. Ratings which are a multiple of 5 are highlighted.

Order Effect	BF_{10}	error%
COE $A1$ $A2$	0.149	9.857×10^{-6}
COE $A1$ $\neg A2$	0.104	1.403×10^{-5}
COE $\neg A1$ $A2$	0.047	3.104×10^{-5}
COE $\neg A1$ $\neg A2$	0.047	3.146×10^{-5}
COE $A2$ $A3$	0.05	2.957×10^{-5}
COE $A2$ $\neg A3$	0.047	3.096×10^{-5}
COE $\neg A2$ $A3$	0.158	9.313×10^{-6}
COE $\neg A2$ $\neg A3$	0.071	2.056×10^{-5}
COE $A3$ $A1$	0.139	1.057×10^{-5}
COE $A3$ $\neg A1$	0.103	1.417×10^{-5}
COE $\neg A3$ $A1$	0.140	1.049×10^{-5}
COE $\neg A3$ $\neg A1$	0.057	2.567×10^{-5}
DOE $A1$ $A2$	0.211	6.929×10^{-6}
DOE $A1$ $\neg A2$	0.120	1.220×10^{-5}
DOE $\neg A1$ $A2$	0.047	3.112×10^{-5}
DOE $\neg A1$ $\neg A2$	0.105	1.389×10^{-5}
DOE $A2$ $A3$	0.161	9.084×10^{-6}
DOE $A2$ $\neg A3$	0.069	2.125×10^{-5}
DOE $\neg A2$ $A3$	0.047	3.127×10^{-5}
DOE $\neg A2$ $\neg A3$	0.050	2.908×10^{-5}
DOE $A3$ $A1$	0.100	1.470×10^{-5}
DOE $A3$ $\neg A1$	0.059	2.499×10^{-5}
DOE $\neg A3$ $A1$	0.057	2.590×10^{-5}
DOE $\neg A3$ $\neg A1$	0.047	3.148×10^{-5}

Table S2.3: Bayesian one sample t-tests of all conjunction order effects (COE) and disjunction order effects (DOE) for Triplet 1, suggesting strong evidence against order effects for all questions pairs. Note: for all tests, the alternative hypothesis specifies that the population mean differs from 0.

Order Effect	BF_{10}	error %
COE $A1$ $A2$	0.233	6.169×10^{-6}
COE $A1$ $\neg A2$	0.059	2.465×10^{-5}
COE $\neg A1$ $A2$	0.120	1.205×10^{-5}
COE $\neg A1$ $\neg A2$	0.047	3.080×10^{-5}
COE $A2$ $A3$	0.115	1.250×10^{-5}
COE $A2$ $\neg A3$	0.051	2.849×10^{-5}
COE $\neg A2$ $A3$	0.051	2.946×10^{-6}
COE $\neg A2$ $\neg A3$	0.047	3.077×10^{-5}
COE $A3$ $A1$	0.054	2.688×10^{-5}
COE $A3$ $\neg A1$	0.051	2.823×10^{-5}
COE $\neg A3$ $A1$	0.114	1.267×10^{-5}
COE $\neg A3$ $\neg A1$	0.063	2.304×10^{-5}
DOE $A1$ $A2$	0.065	2.234×10^{-5}
DOE $A1$ $\neg A2$	0.052	2.794×10^{-5}
DOE $\neg A1$ $A2$	0.060	2.423×10^{-5}
DOE $\neg A1$ $\neg A2$	0.051	2.830×10^{-5}
DOE $A2$ $A3$	0.143	1.010×10^{-5}
DOE $A2$ $\neg A3$	0.052	2.782×10^{-5}
DOE $\neg A2$ $A3$	0.047	3.076×10^{-5}
DOE $\neg A2$ $\neg A3$	0.052	2.785×10^{-5}
DOE $A3$ $A1$	0.050	2.910×10^{-5}
DOE $A3$ $\neg A1$	$\mathbf{3.049 \times 10^{7}}$	$\mathbf{4.366 \times 10^{-14}}$
DOE $\neg A3$ $A1$	0.097	1.486×10^{-5}
DOE $\neg A3$ $\neg A1$	0.051	2.850×10^{-5}

Table S2.4: Bayesian one sample t-tests of all conjunction order effects (COE) and disjunction order effects (DOE) for Triplet 2, suggesting strong evidence against order effects for all but one (printed in boldface) question pair. Note: for all tests, the alternative hypothesis specifies that the population mean differs from 0.

Measure 1	Measure 2	BF_{01} triplet 1	error%	BF_{01} triplet 2	error%
$P(A1\|A2)$	$P(A2\|A1)$	0.016	2.385×10^{-8}	0.010	1.436×10^{-8}
$P(A1\|\neg A2)$	$P(\neg A2\|A1)$	**10.431**	1.528×10^{-5}	1.085×10^{-5}	1.436×10^{-14}
$P(\neg A1\|A2)$	$P(A2\|\neg A1)$	**15.018**	2.202×10^{-5}	1.851×10^{-9}	2.426×10^{-15}
$P(\neg A1\|\neg A2)$	$P(\neg A2\|\neg A1)$	4.028×10^{-6}	5.671×10^{-12}	1.434×10^{-4}	1.989×10^{-10}
$P(A2\|A3)$	$P(A3\|A2)$	1.487×10^{-12}	1.952×10^{-18}	9.055×10^{-18}	1.033×10^{-23}
$P(A2\|\neg A3)$	$P(\neg A3\|A2)$	**3.658**	5.349×10^{-6}	0.871	1.248×10^{-6}
$P(\neg A2\|A3)$	$P(A3\|\neg A2)$	**1.884**	2.757×10^{-6}	6.104×10^{-4}	8.517×10^{-10}
$P(\neg A2\|\neg A3)$	$P(\neg A3\|\neg A2)$	2.888×10^{-13}	3.755×10^{-19}	1.658×10^{-21}	1.749×10^{-27}
$P(A3\|A1)$	$P(A1\|A3)$	4.763×10^{-5}	6.767×10^{-11}	1.477×10^{-9}	1.932×10^{-15}
$P(A3\|\neg A1)$	$P(\neg A1\|A3)$	0.054	7.859×10^{-8}	**12.678**	1.829×10^{-5}
$P(\neg A3\|A1)$	$P(A1\|\neg A3)$	0.003	3.677×10^{-9}	**7.378**	1.063×10^{-5}
$P(\neg A3\|\neg A1)$	$P(\neg A1\|\neg A3)$	2.581×10^{-6}	3.623×10^{-12}	6.349×10^{-10}	8.262×10^{-16}

Table S2.5: Bayesian paired sample t-tests comparing conditional probabilities to their reciprocal counterpart, for both triplets together. Results confirming the assumption of reciprocity are printed in boldface.

	BF_{01}	error%		BF_{01}	error%
$\boldsymbol{Z_1 A1, A2}$	**1.199**	**0.000**	$\boldsymbol{Z_9 A1, A2}$	**27.306**	**0.011**
$\boldsymbol{Z_1 A2, A1}$	**13.706**	**0.006**	$\boldsymbol{Z_9 A2, A3}$	**9.120**	**0.004**
$\boldsymbol{Z_1 A2, A3}$	**11.802**	**0.005**	$\boldsymbol{Z_9 A3, A1}$	**15.034**	**0.006**
$\boldsymbol{Z_1 A3, A2}$	**24.521**	**0.010**	$Z_{10}A1, A2$	0.000	0.000
$\boldsymbol{Z_1 A3, A1}$	**1.271**	**0.000**	$Z_{10}A2, A3$	0.000	0.000
$\boldsymbol{Z_1 A1, A3}$	**1.637**	**0.000**	$Z_{10}A3, A1$	0.000	0.000
$\boldsymbol{Z_2 A1, A2}$	**26.892**	**0.011**	$Z_{11}A1, A2$	0.000	0.000
$\boldsymbol{Z_2 A2, A1}$	**19.309**	**0.008**	$Z_{11}A2, A3$	0.000	0.000
$\boldsymbol{Z_2 A2, A3}$	**30.067**	**0.012**	$Z_{11}A3, A1$	0.000	0.000
$\boldsymbol{Z_2 A3, A2}$	**15.176**	**0.006**	$Z_{12}A1, A2$	0.000	0.000
$\boldsymbol{Z_2 A3, A1}$	**21.341**	**0.009**	$Z_{12}A2, A3$	0.000	0.000
$\boldsymbol{Z_2 A1, A3}$	**29.938**	**0.012**	$Z_{12}A3, A1$	0.000	0.000
$Z_3A1, A2$	0.000	0.000	$Z_{13}A1, A2$	0.000	0.000
$Z_3A2, A1$	0.000	0.000	$Z_{13}A2, A3$	0.000	0.000
$Z_3A2, A3$	0.000	0.000	$Z_{13}A3, A1$	0.000	0.000
$Z_3A3, A2$	0.000	0.000	$\boldsymbol{Z_{14} A1 A2}$	**27.045**	**0.011**
$Z_3A3, A1$	0.000	0.000	$\boldsymbol{Z_{14} A2, A3}$	**30.189**	**0.012**
$Z_3A1, A3$	0.000	0.000	$\boldsymbol{Z_{14} A3, A1}$	**25.437**	**0.010**
$Z_4A1, A2$	0.000	0.000	$Z_{15}A1, A2$	0.000	0.000
$Z_4A2, A1$	0.000	0.000	$Z_{15}A2, A1$	0.000	0.000
$Z_4A2, A3$	0.000	0.000	$Z_{15}A2, A3$	0.000	0.000
$Z_4A3, A2$	0.000	0.000	$Z_{15}A3, A2$	0.000	0.000
$Z_4A3, A1$	0.000	0.000	$Z_{15}A3, A1$	0.000	0.000
$Z_4A1, A3$	0.000	0.000	$Z_{15}A1, A3$	0.000	0.000

$Z_6A1, A2$	0.000	0.000	$Z_{17}A1, A2$	0.000	0.000
$Z_6A2, A1$	0.000	0.000	$Z_{17}A2, A1$	0.000	0.000
$Z_6A2, A3$	0.000	0.000	$Z_{17}A2, A3$	0.000	0.000
$Z_6A3, A2$	0.000	0.000	$Z_{17}A3, A2$	0.000	0.000
$Z_6A3, A1$	0.000	0.000	$Z_{17}A3, A1$	0.000	0.000
$Z_6A1, A3$	0.000	0.000	$Z_{17}A1, A3$	0.000	0.000
$Z_7A1, A2$	0.000	0.000	$Z_{18}A1, A2$	0.000	0.000
$Z_7A2, A1$	0.000	0.000	$Z_{18}A2, A1$	0.000	0.000
$Z_7A2, A3$	0.000	0.000	$Z_{18}A2, A3$	0.000	0.000
$Z_7A3, A2$	0.000	0.000	$Z_{18}A3, A2$	0.000	0.000
$Z_7A3, A1$	0.000	0.000	$Z_{18}A3, A1$	0.000	0.000
$Z_7A1, A3$	0.000	0.000	$Z_{18}A1, A3$	0.000	0.000
$Z_8A1, A2$	0.000	0.000			
$Z_8A2, A1$	0.000	0.000			
$Z_8A2, A3$	0.000	0.000			
$Z_8A3, A2$	0.000	0.000			
$Z_8A3, A1$	0.000	0.000			
$Z_8A1, A3$	0.000	0.000			

Table S2.6: Bayesian one sample t-tests for all the Z identities, for both triplets together (Table 2 in main text). Results that confirm Bayesian probability theory are printed in boldface. Note, we evaluated the identities separately for each order (for conjunctions, disjunctions). Note: for all tests, the null hypothesis specifies that the population mean equals 0. Figure 4 in main text illustrates the results differently, separating them out by triplet; it can be seen that there is hardly any difference between the two triplets, hence the aggregated presentation in this table.

	BF_{10}	error %
$T1\ P(A1) + P(\neg A1)$	1.216×10^{56}	8.807×10^{-59}
$T1\ P(A2) + P(\neg A2)$	1.020×10^{56}	9.258×10^{-60}
$T1\ P(A3) + P(\neg A3)$	6.273×10^{53}	5.756×10^{-61}
$T1\ P(A1 \wedge A2) + P(A1 \wedge \neg A2) + P(\neg A1 \wedge A2) + P(\neg A1 \wedge \neg A2)$	1.157×10^{184}	4.970×10^{-188}
$T1\ P(A2 \wedge A3) + P(A2 \wedge \neg A3) + P(\neg A2 \wedge A3) + P(\neg A2 \wedge \neg A3)$	1.216×10^{201}	1.546×10^{-250}
$T1\ P(A3 \wedge A1) + P(A3 \wedge \neg A1) + P(\neg A3 \wedge A1) + P(\neg A3 \wedge \neg A1)$	5.853×10^{190}	1.547×10^{-195}
$T1\ P(A2 \wedge A1) + P(A2 \wedge \neg A1) + P(\neg A2 \wedge A1) + P(\neg A2 \wedge \neg A1)$	4.127×10^{204}	4.930×10^{-209}
$T1\ P(A3 \wedge A2) + P(A3 \wedge \neg A2) + P(\neg A3 \wedge A2) + P(\neg A3 \wedge \neg A2)$	1.192×10^{188}	6.566×10^{-193}
$T1\ P(A1 \wedge A3) + P(A1 \wedge \neg A3) + P(\neg A1 \wedge A3) + P(\neg A1 \wedge \neg A3)$	2.036×10^{189}	4.822×10^{-194}
$T2\ P(A1) + P(\neg A1)$	1.186×10^{53}	8.359×10^{-56}
$T2\ P(A2) + P(\neg A2)$	3.444×10^{50}	1.032×10^{-57}
$T2\ P(A3) + P(\neg A3)$	5.499×10^{50}	6.395×10^{-58}
$T2\ P(A1 \wedge A2) + P(A1 \wedge \neg A2) + P(\neg A1 \wedge A2) + P(\neg A1 \wedge \neg A2)$	3.766×10^{167}	2.172×10^{-170}
$T2\ P(A1 \wedge A2) + P(A1 \wedge \neg A2) + P(\neg A1 \wedge A2) + P(\neg A1 \wedge \neg A2)$	3.766×10^{167}	2.172×10^{-170}
$T2\ P(A2 \wedge A3) + P(A2 \wedge \neg A3) + P(\neg A2 \wedge A3) + P(\neg A2 \wedge \neg A3)$	1.132×10^{171}	2.294×10^{-174}
$T2\ P(A3 \wedge A1) + P(A3 \wedge \neg A1) + P(\neg A3 \wedge A1) + P(\neg A3 \wedge \neg A1)$	8.408×10^{182}	1.025×10^{-186}
$T2\ P(A2 \wedge A1) + P(A2 \wedge \neg A1) + P(\neg A2 \wedge A1) + P(\neg A2 \wedge \neg A1)$	3.706×10^{164}	6.860×10^{-168}
$T2\ P(A3 \wedge A2) + P(A3 \wedge \neg A2) + P(\neg A3 \wedge A2) + P(\neg A3 \wedge \neg A2)$	9.390×10^{165}	4.566×10^{-169}
$T2\ P(A1 \wedge A3) + P(A1 \wedge \neg A3) + P(\neg A1 \wedge A3) + P(\neg A1 \wedge \neg A3)$	5.264×10^{-178}	9.444×10^{-182}

Table S2.7: One-sample Bayesian t-tests concerning consistency with the binary complementarity on marginals and four-way law of total probability. Note: For all tests, the null hypothesis is that the population mean equals 1. T1 and T2 refer to the two different triplets.

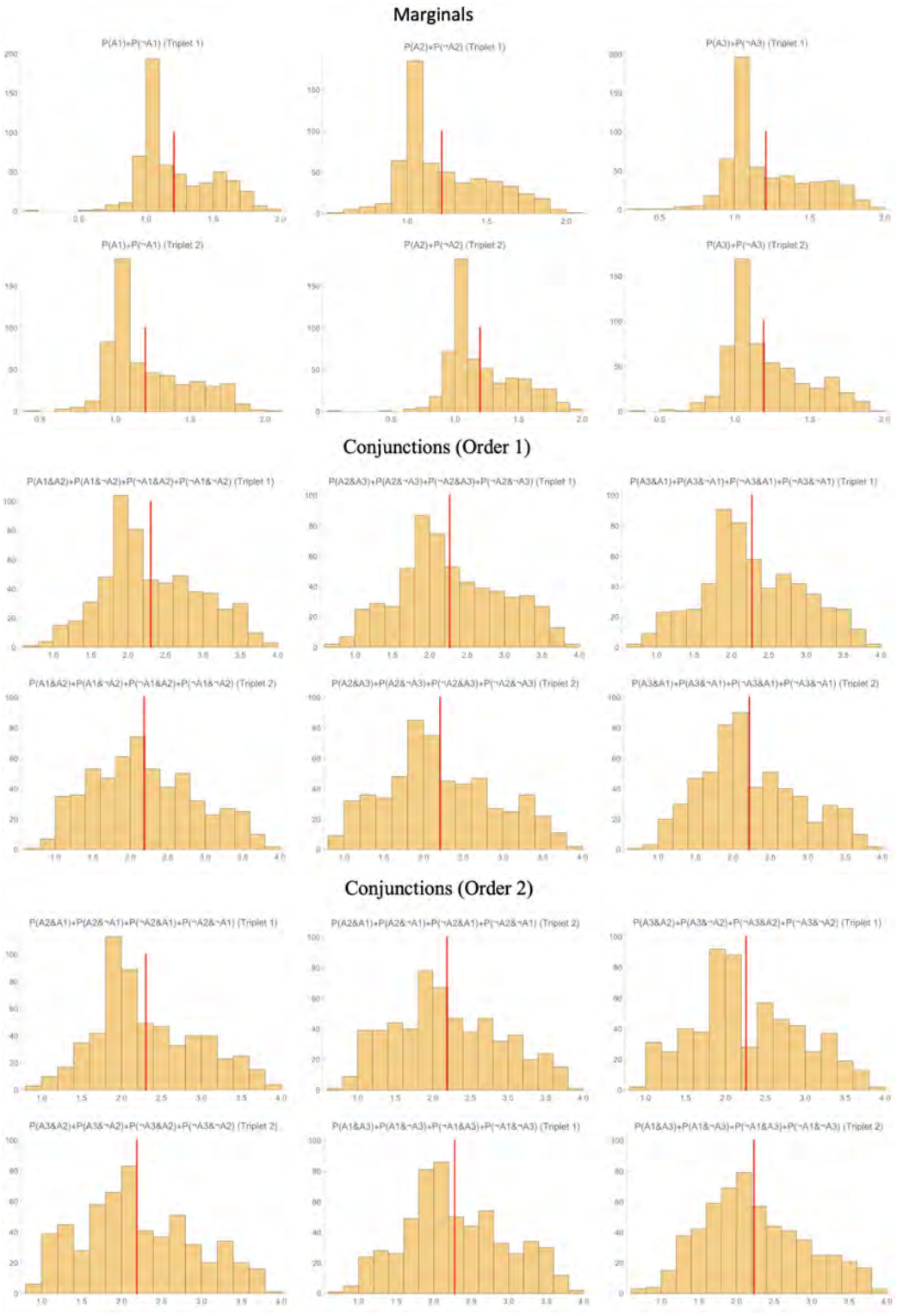
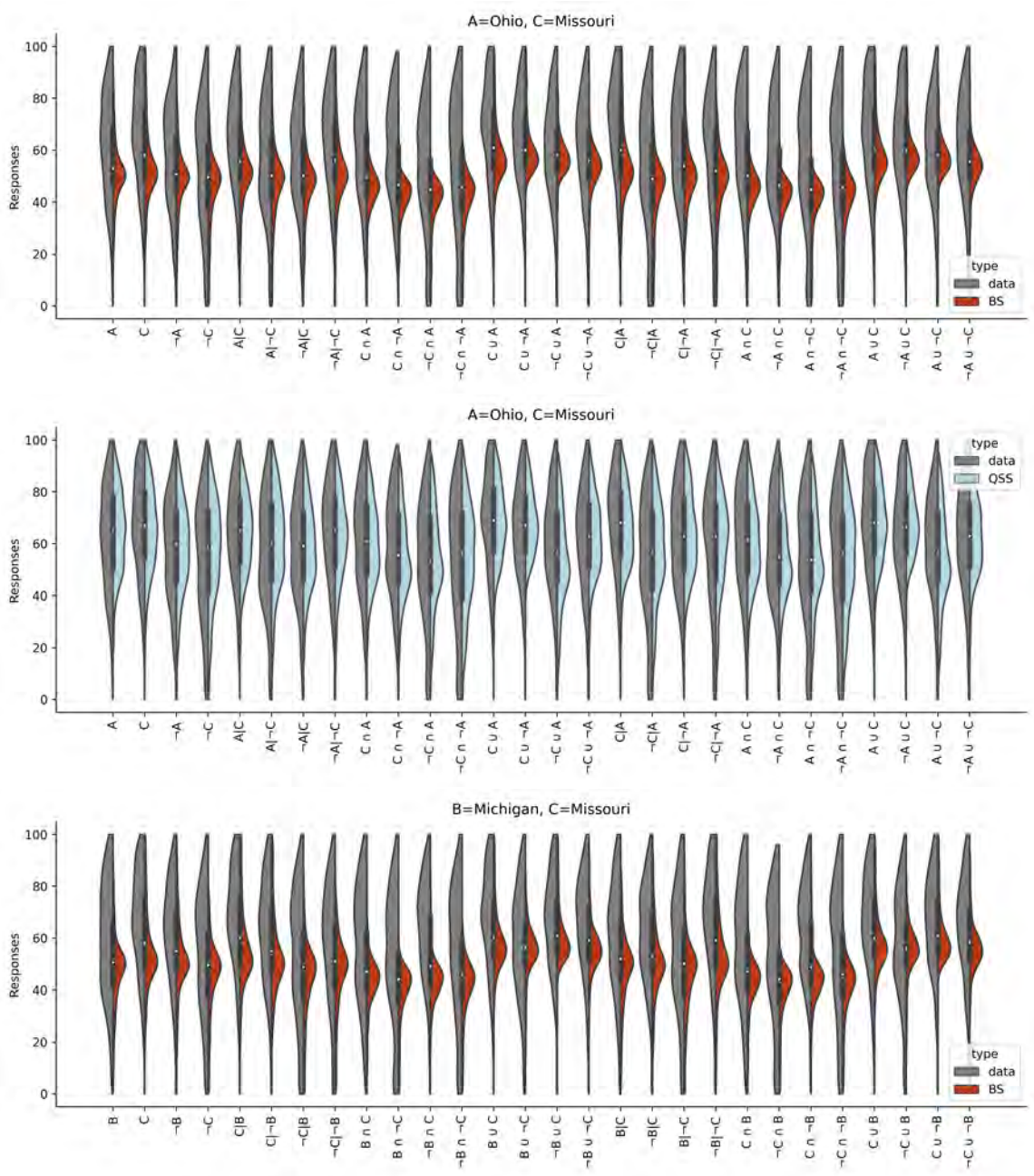


Figure S.2.2. Histograms of the quantities corresponding to binary complementarity on marginals and the four-way law of total probability, for different triplets and for different predicate orders for conjunctions. In all cases, the observed values exceed the expected value of 1.

Supplementary Material 3

In this section, we offer some additional figures, illustrating the results of the comparison between the Quantum Sequential Sampler and the Bayesian Sampler models.



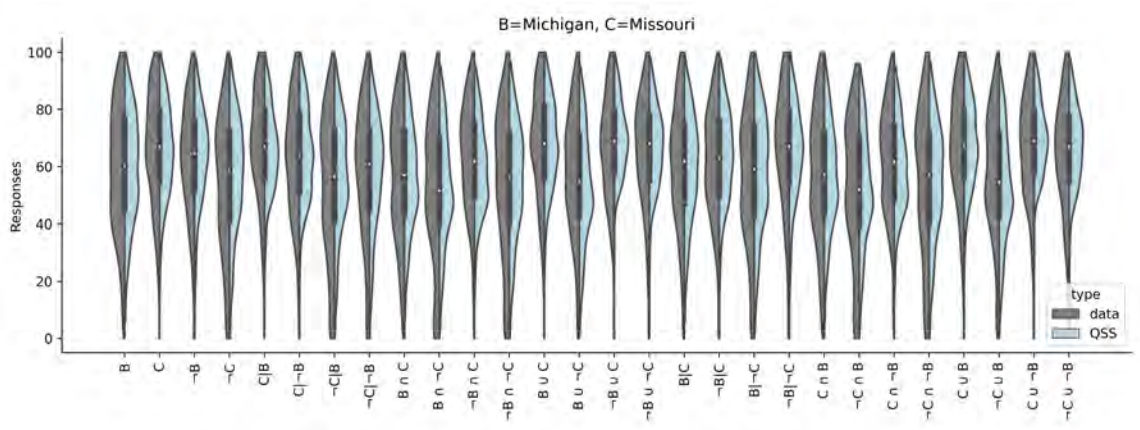


Figure S.3.1. Violin plots showing empirical data vs. the predictions of the Quantum Sequential Sampler, and empirical data (top panel) vs. the predictions of the Bayesian Sampler (bottom panel) for the first triplet. The data are for the $\{A, C\}$ and $\{B, C\}$ pairs of events of Triplet 1.

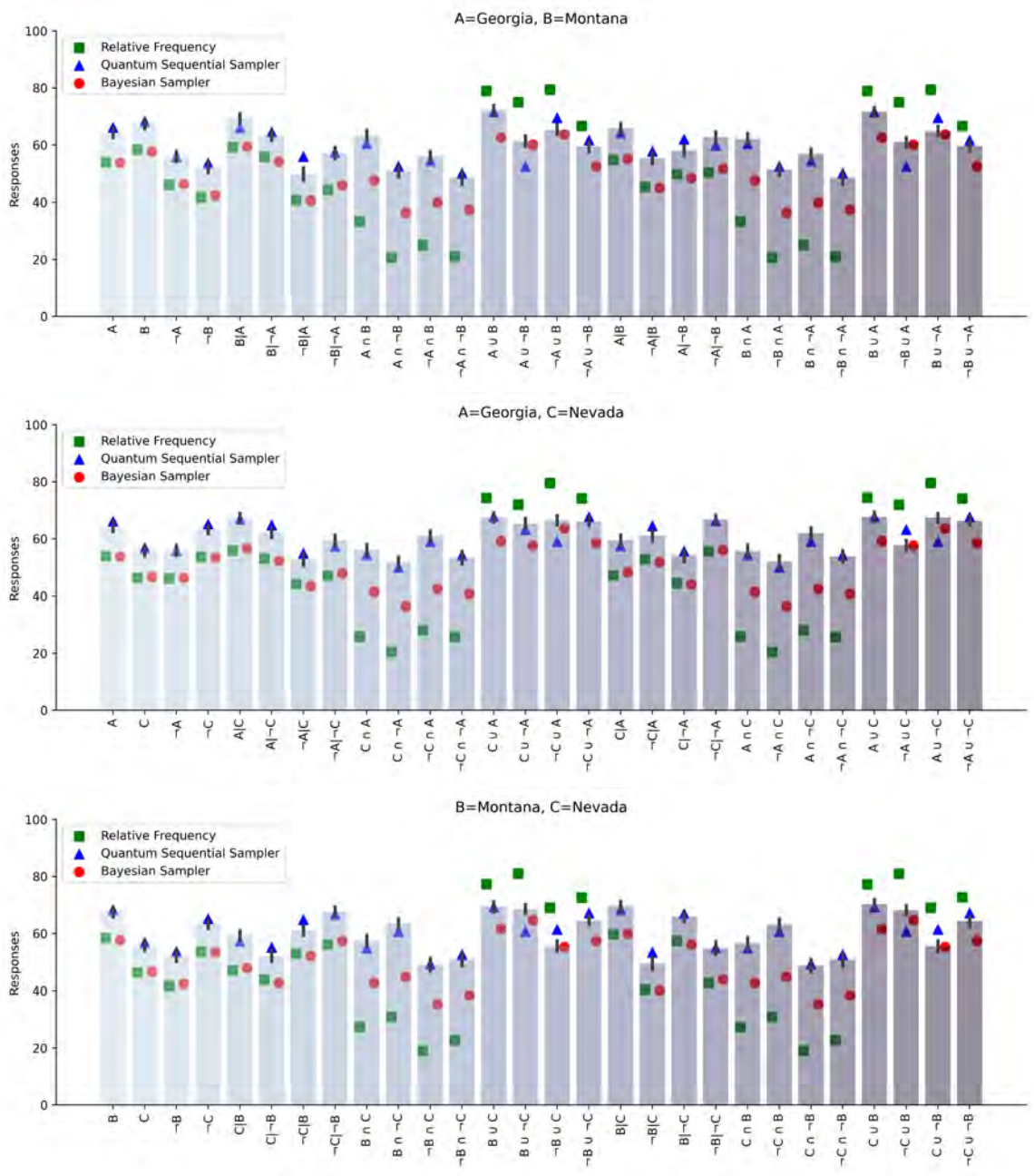
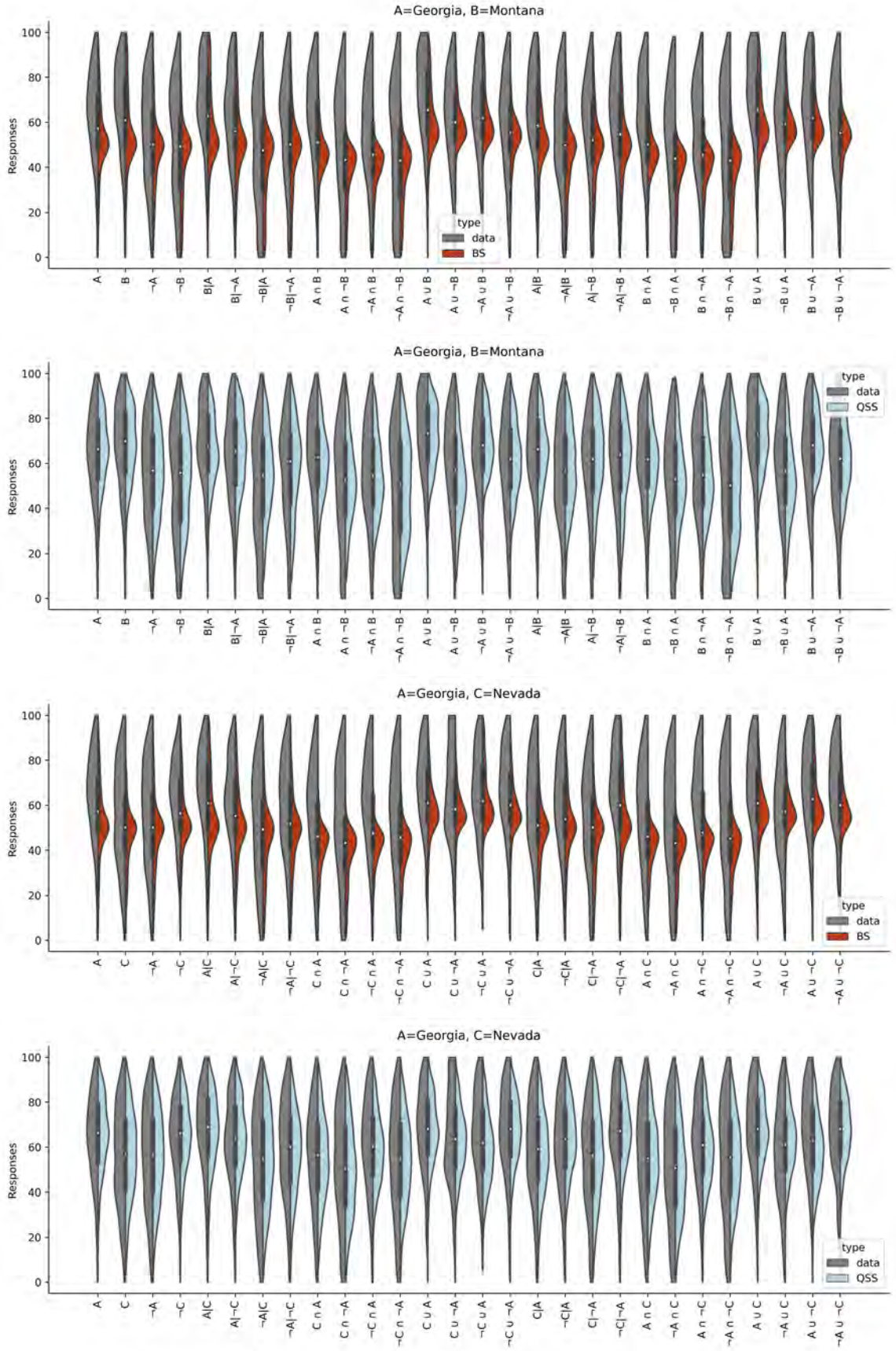


Figure S.3.2. Mean predictions of the Quantum Sequential Sampler, Bayesian Sampler, and relative frequency models, against empirical results. The data are for Triplet 2.


A=Georgia, B=Montana
Responses
type
data
BS
A=Georgia, B=Montana
Responses
type
data
QSS
A=Georgia, C=Nevada
Responses
type
data
BS
A=Georgia, C=Nevada
Responses
type
data
QSS

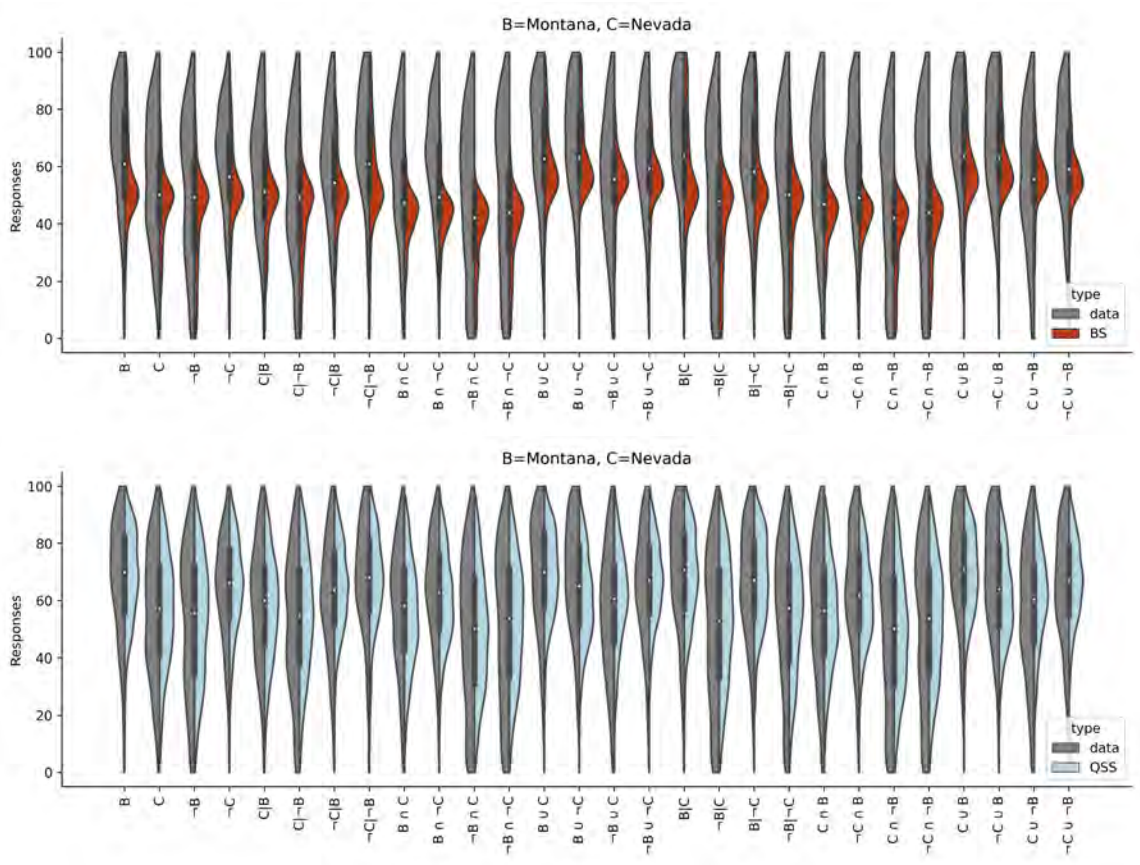


Figure S.3.3. Violin plots showing empirical data vs. the predictions of the Quantum Sequential Sampler, and empirical data (top panel) vs. the predictions of the Bayesian Sampler (bottom panel) for the second triplet. The data are for the $\{A, B\}$, $\{A, C\}$ and $\{B, C\}$ pairs of events for Triplet 2.

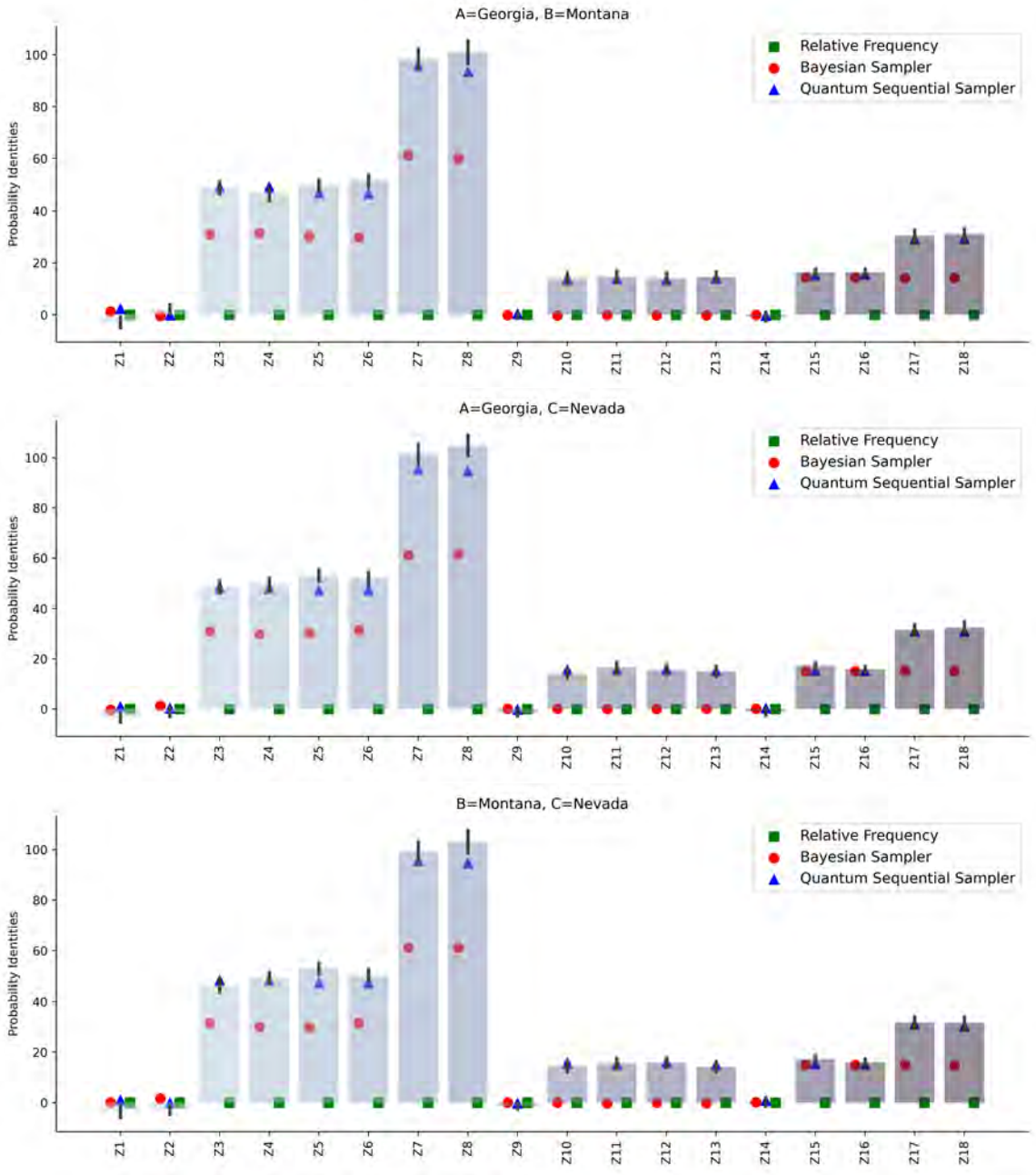


Figure S.3.4. The empirical value of probabilistic identities in Table 2, together with values computed from best-fit predictions, from the Quantum, Sequential Sampler, the Bayesian Sampler, and the relative frequency model. The data in this figure is from Triplet 2.

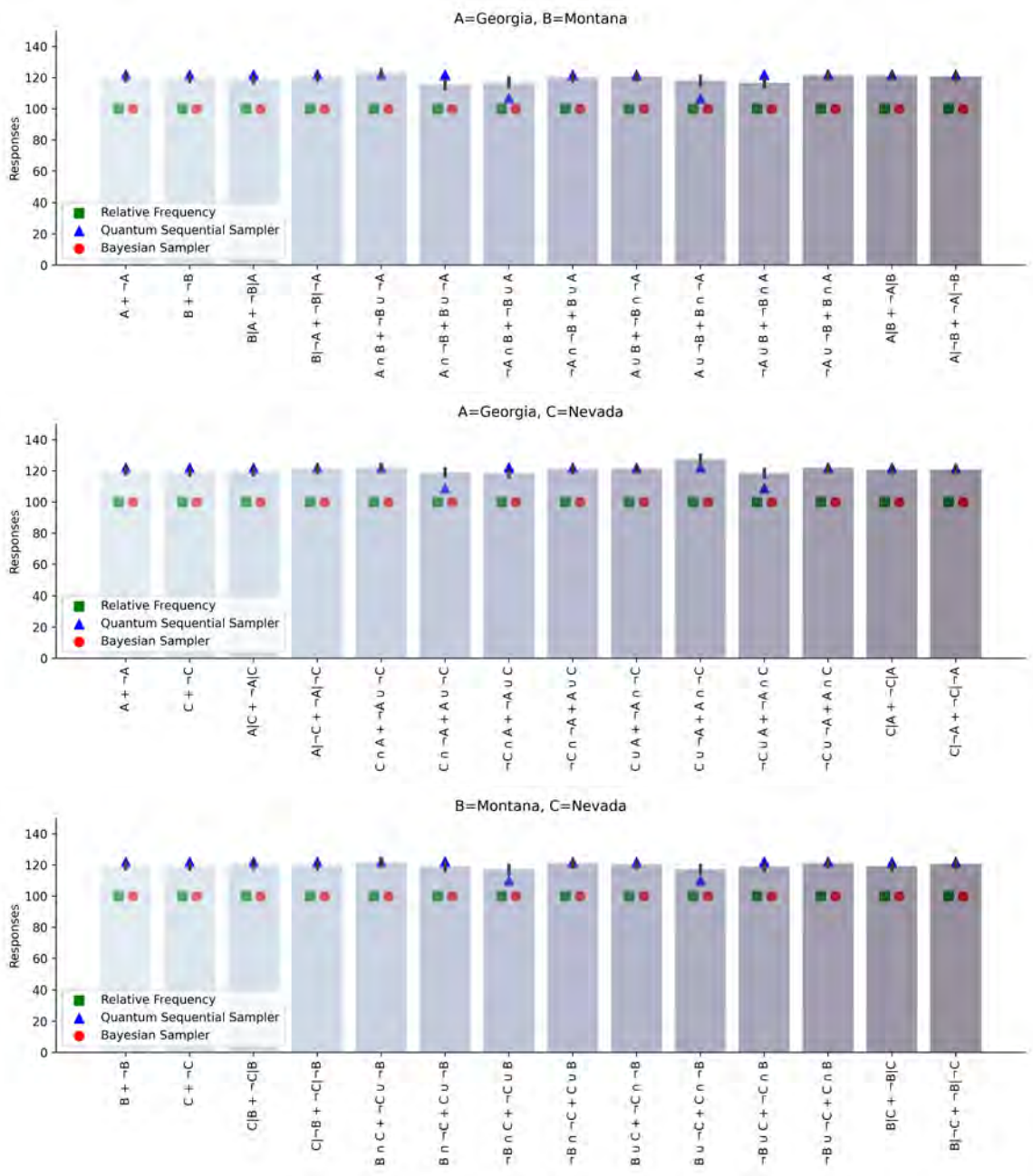
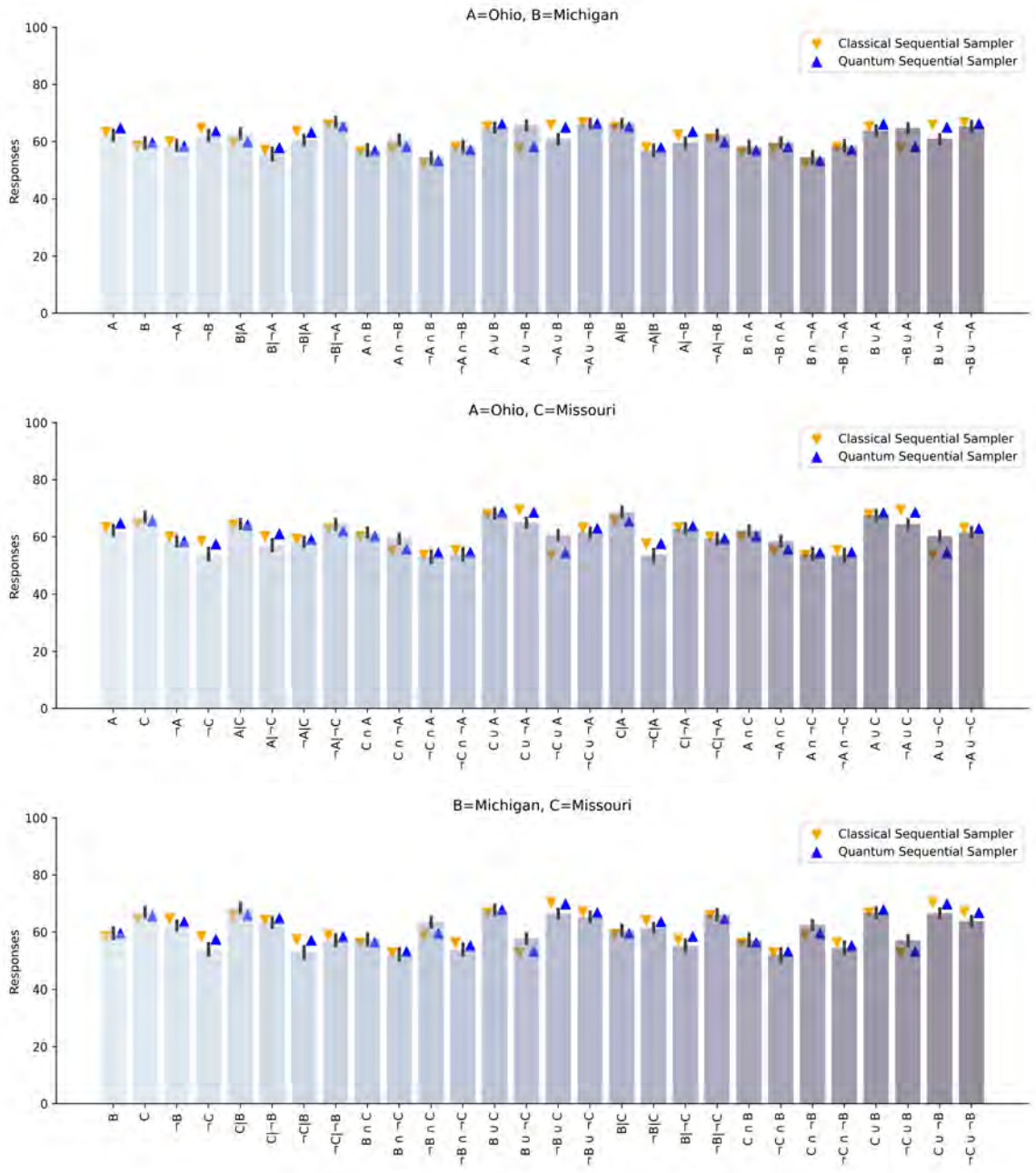


Figure S.3.5. Empirical values for binary complementarity, together with predictions from the Quantum Sequential Sampler, the Bayesian Sampler, and the relative frequency model, averaged across all participants. The results are for Triplet $T2$.

Supplementary Material 4

In this section we offer some additional figures, illustrating the results of comparing the quantum and classical variant of Quantum Sequential Sampler, across all of the 1162 participants.



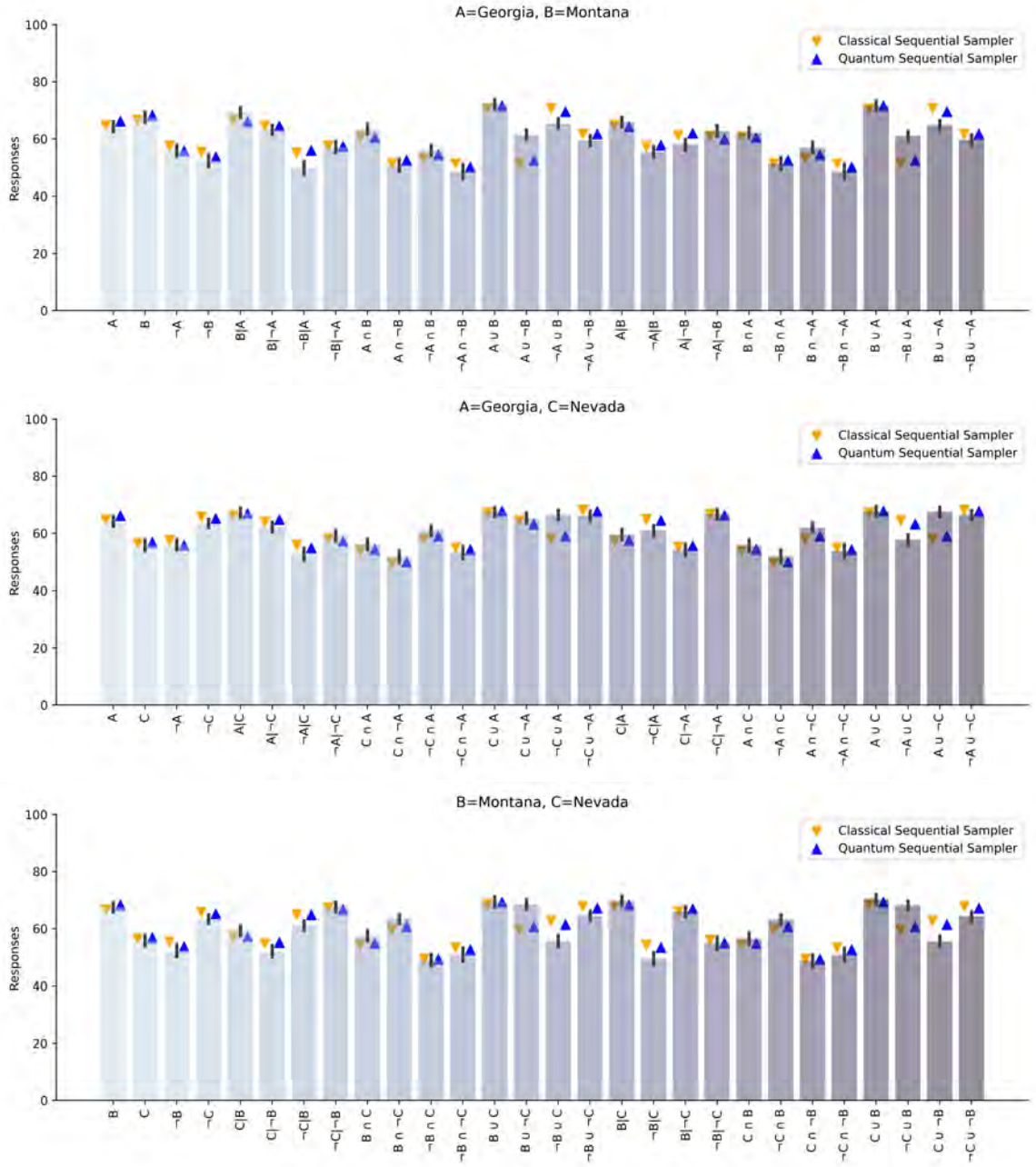
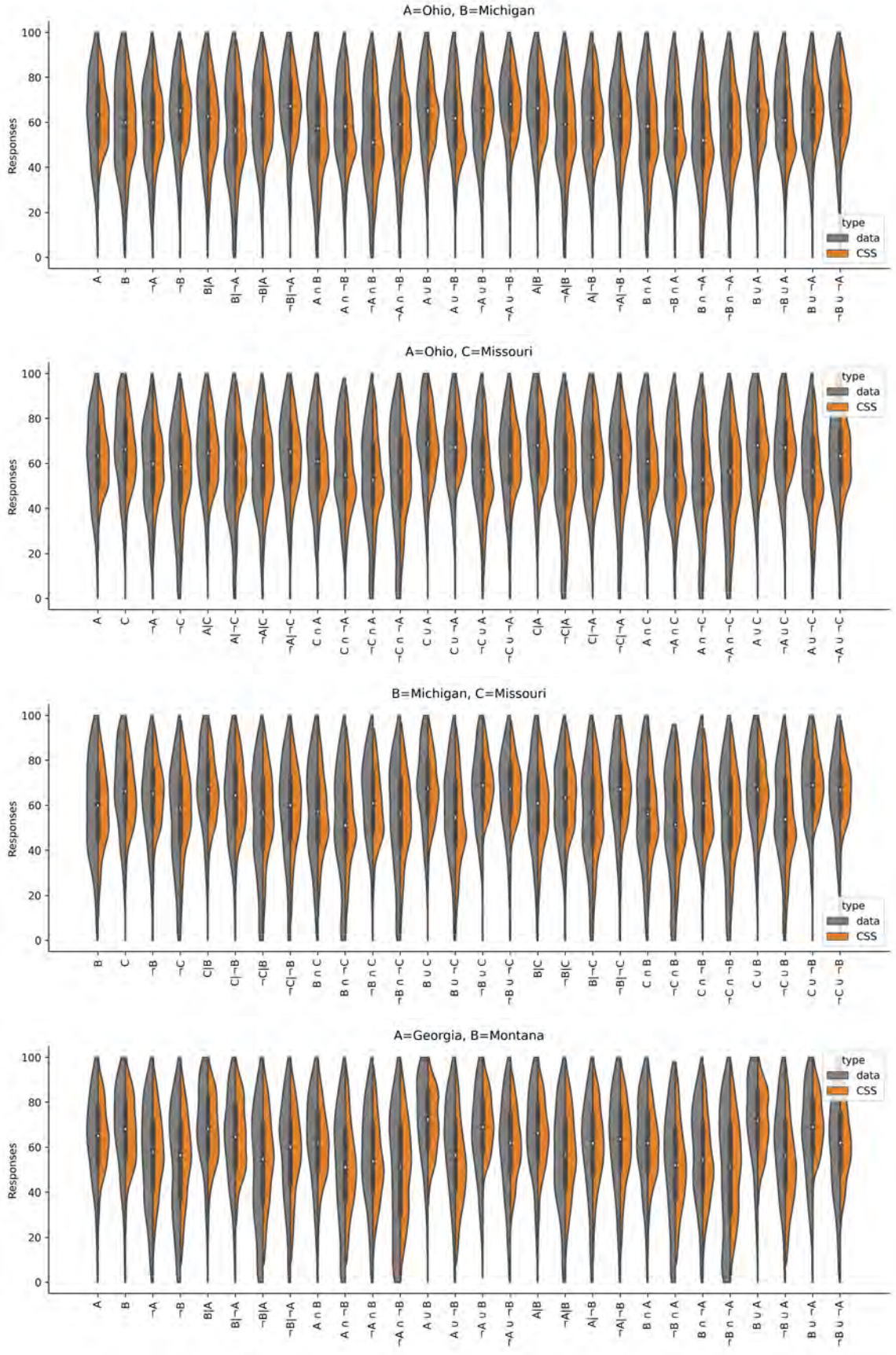


Figure S.4.1. Mean predictions of the quantum and classical variants of the Quantum Sequential Sampler against empirical results.


A=Ohio, B=Michigan
A=Ohio, C=Missouri
B=Michigan, C=Missouri
A=Georgia, B=Montana
Responses
type
data
CSS

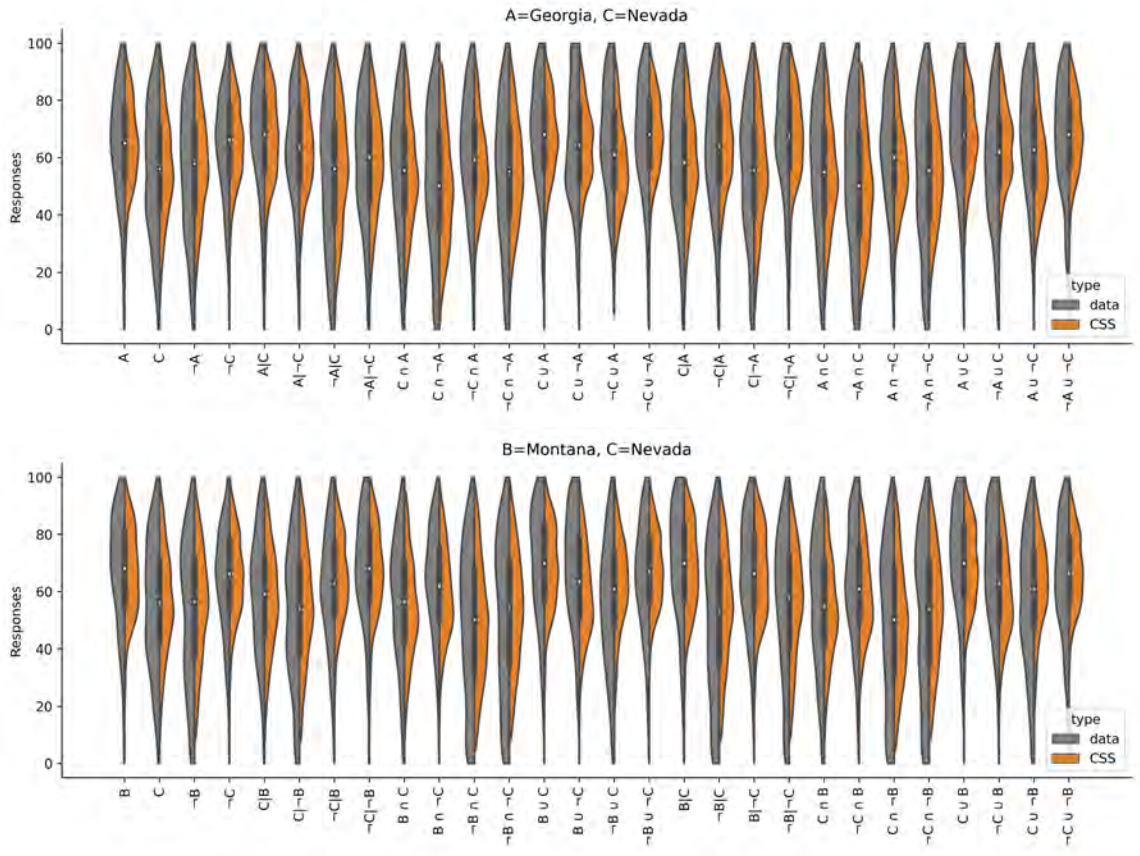
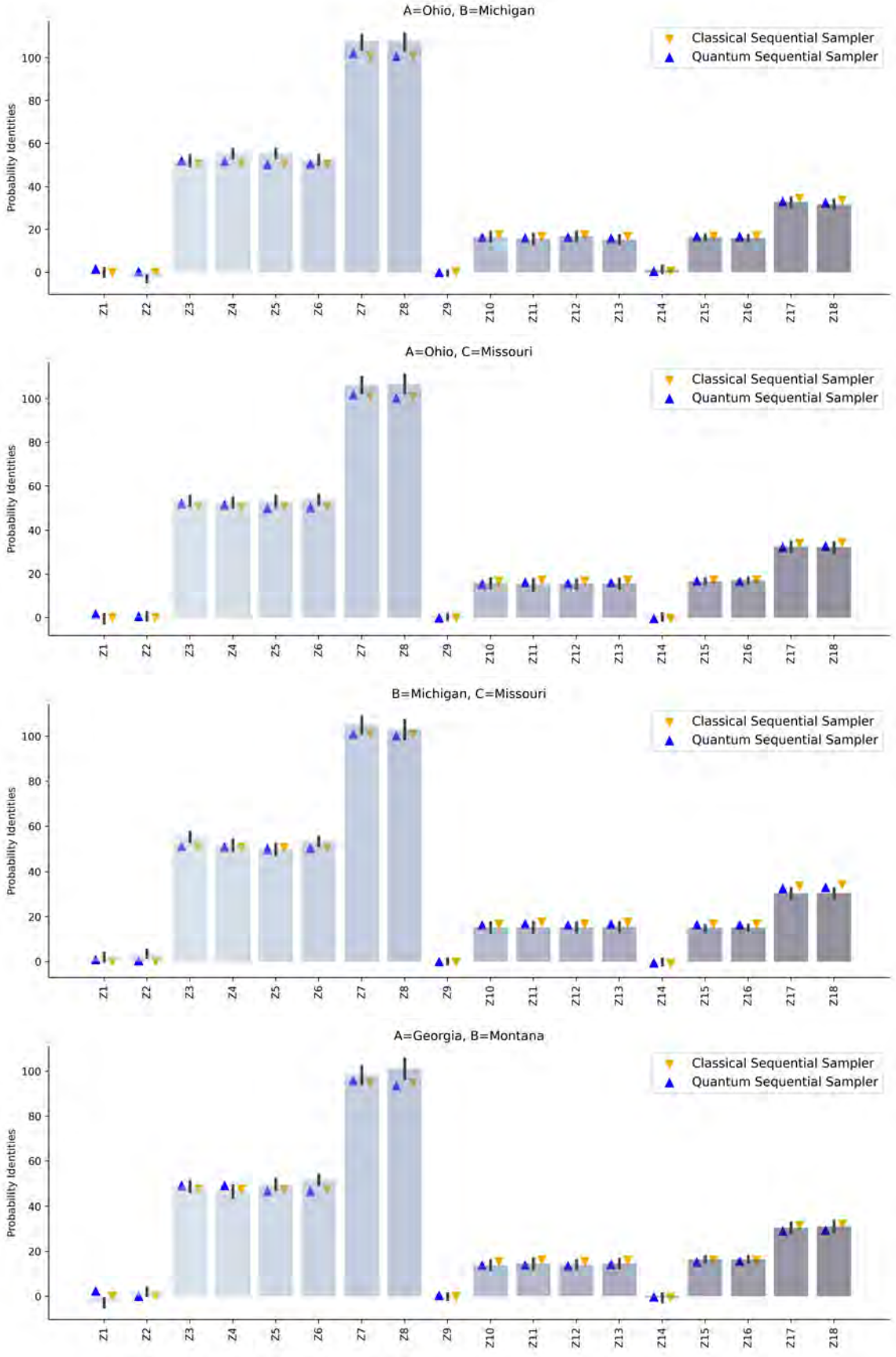


Figure S.4.2. Violin plots showing empirical data vs. the predictions of the classical variant of the Quantum Sequential Sampler model.


A=Ohio, B=Michigan
A=Ohio, C=Missouri
B=Michigan, C=Missouri
A=Georgia, B=Montana
Classical Sequential Sampler
Quantum Sequential Sampler
Probability Identities
0
20
40
60
80
100
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Z9
Z10
Z11
Z12
Z13
Z14
Z15
Z16
Z17
Z18

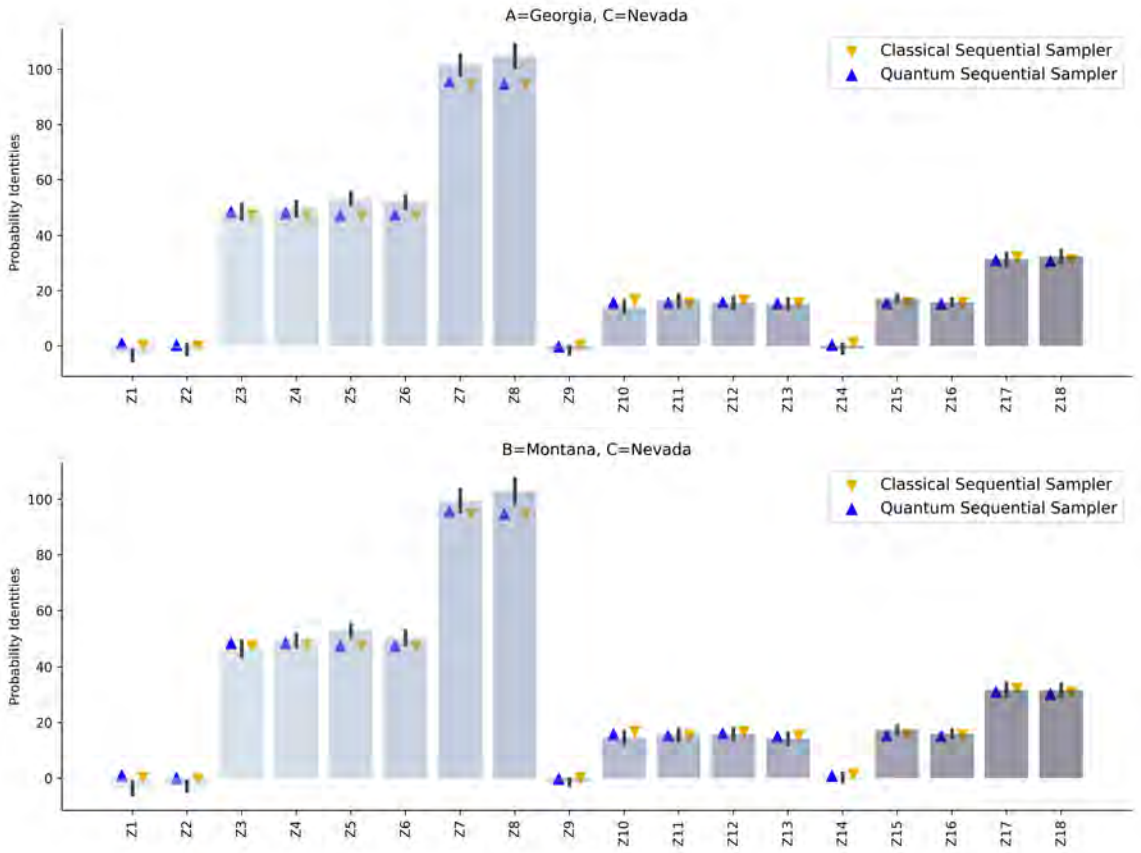
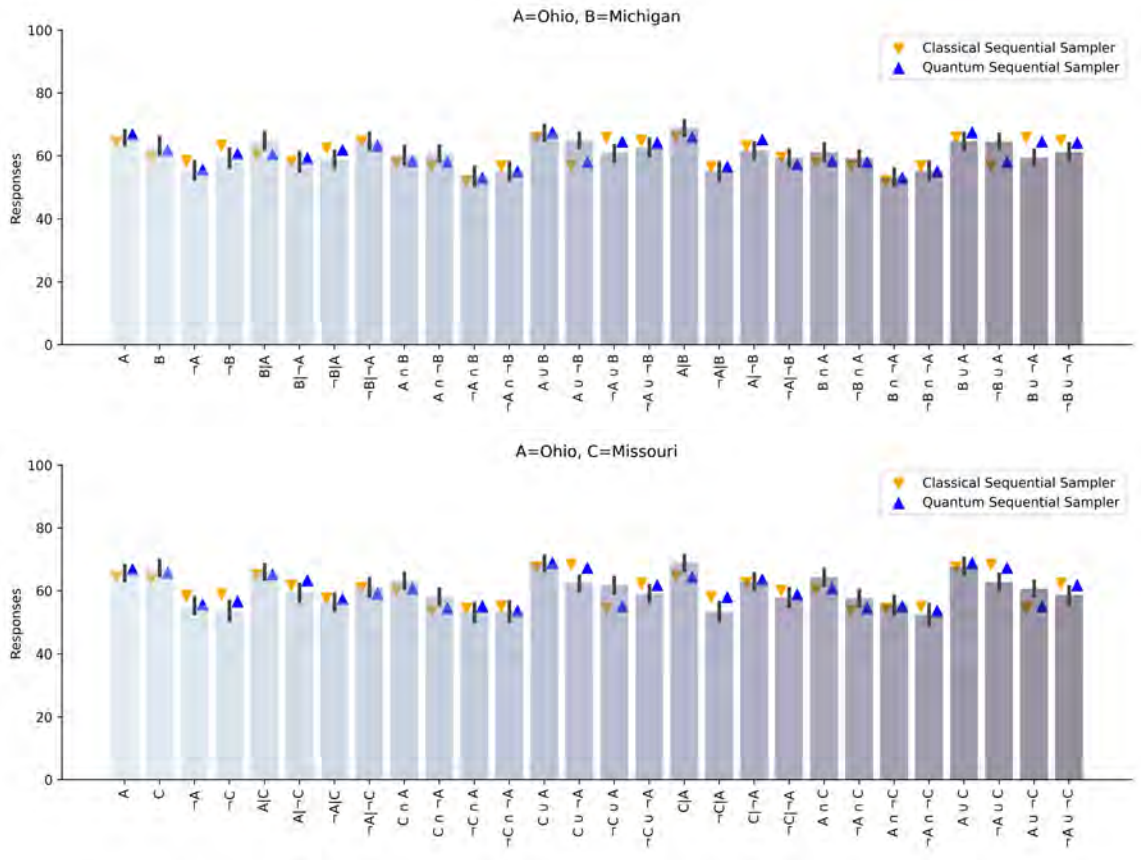
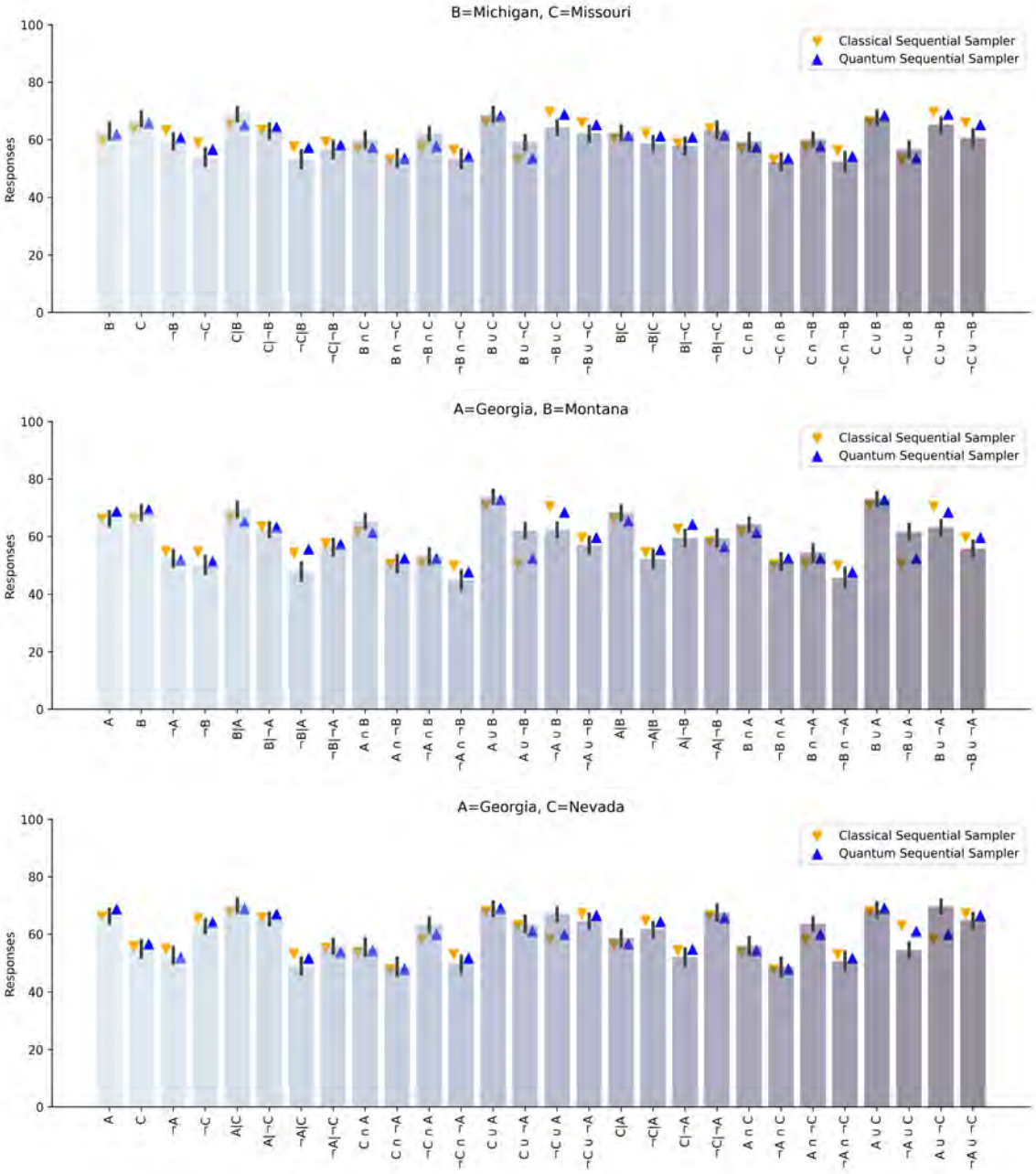


Figure S.4.3. The empirical value of probabilistic identities in Table 2, together with values computed from best-fit predictions, from the quantum and classical variants of the Quantum Sequential Sampler.

Supplementary Material 5

In this section, we display additional figures comparing participants whose quantum interference parameters are significant by the two-sigma criterion ($p < .05$) with those whose quantum interference parameters are significant by the four-sigma criterion ($p < .00008$).




B=Michigan, C=Missouri
A=Georgia, B=Montana
A=Georgia, C=Nevada
Classical Sequential Sampler
Quantum Sequential Sampler
Responses
0
20
40
60
80
100

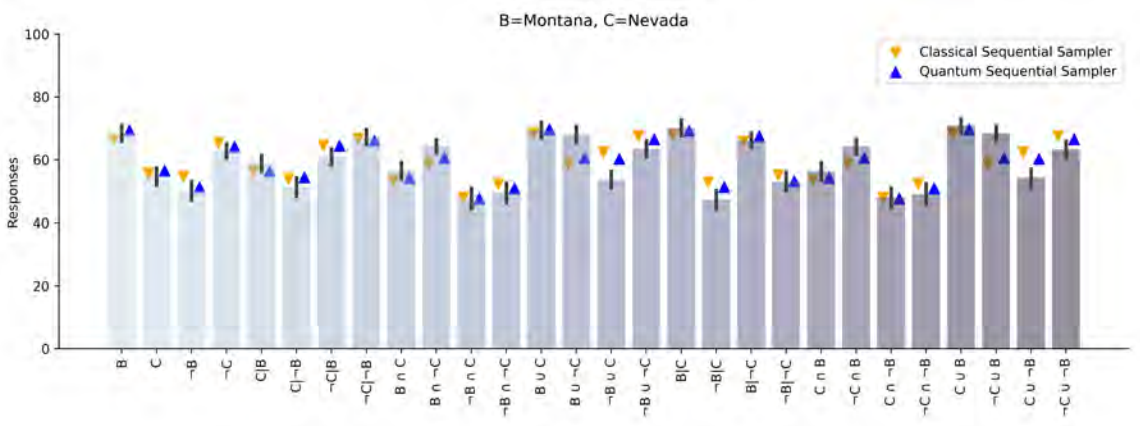
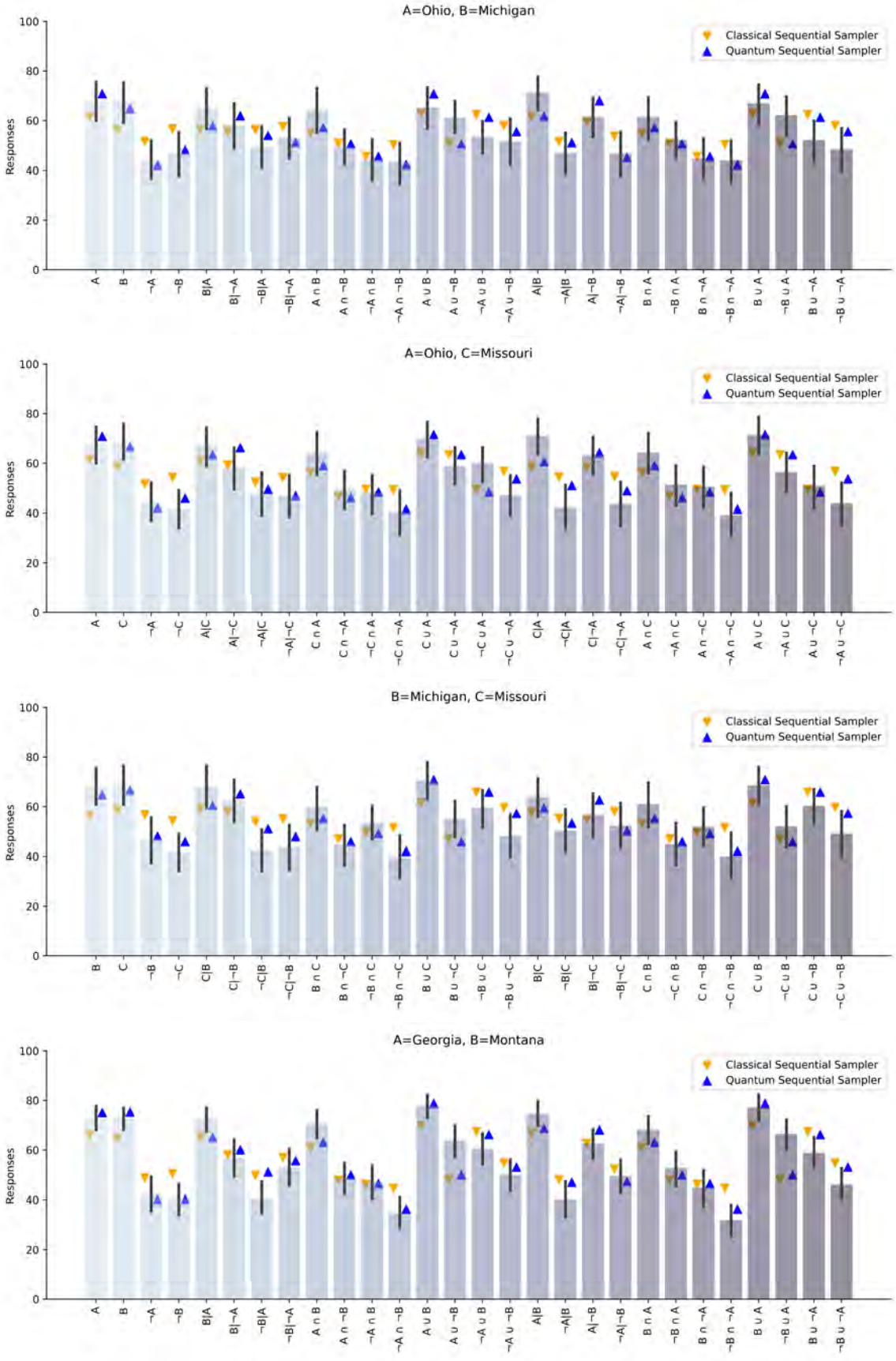


Figure S.5.1. Mean predictions of the quantum and classical variants of the Quantum Sequential Sampler against empirical results for participants whose quantum interference parameter is significant by the criterion ($p < .05$).


A=Ohio, B=Michigan
A=Ohio, C=Missouri
B=Michigan, C=Missouri
A=Georgia, B=Montana
Classical Sequential Sampler
Quantum Sequential Sampler
Responses
0
20
40
60
80
100

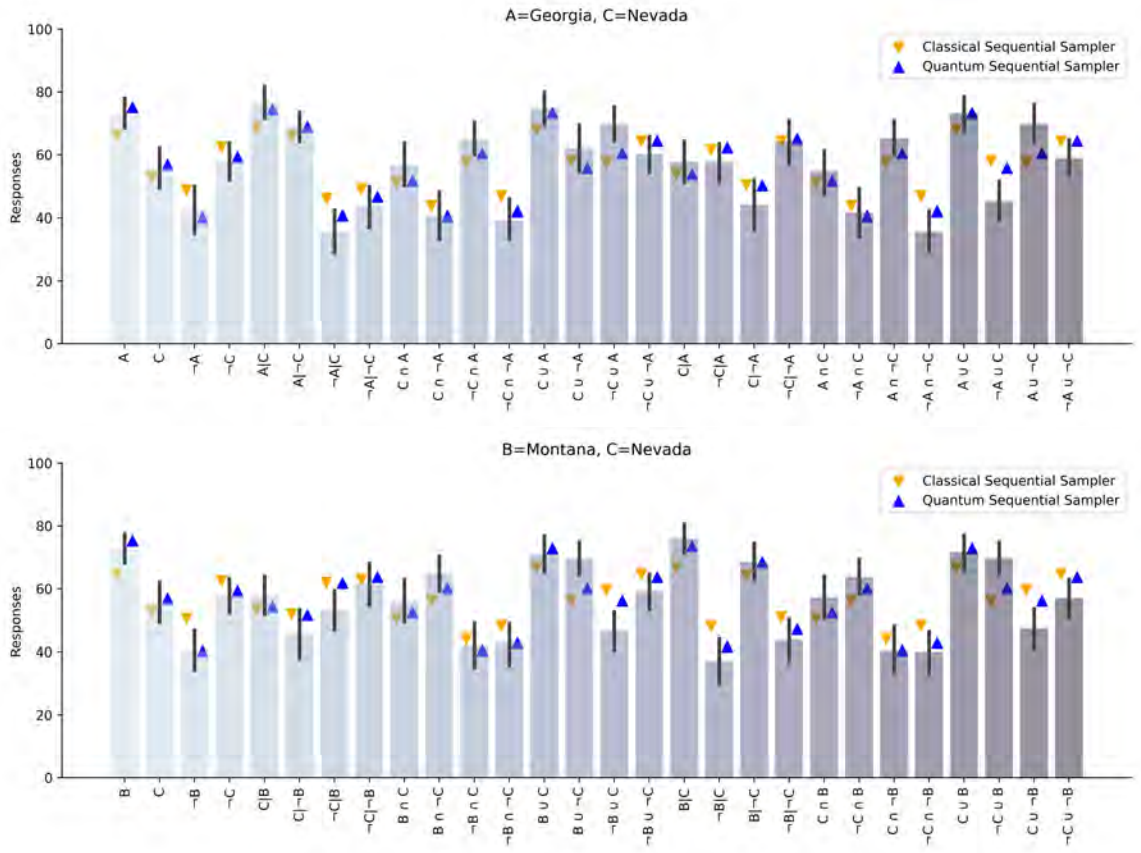
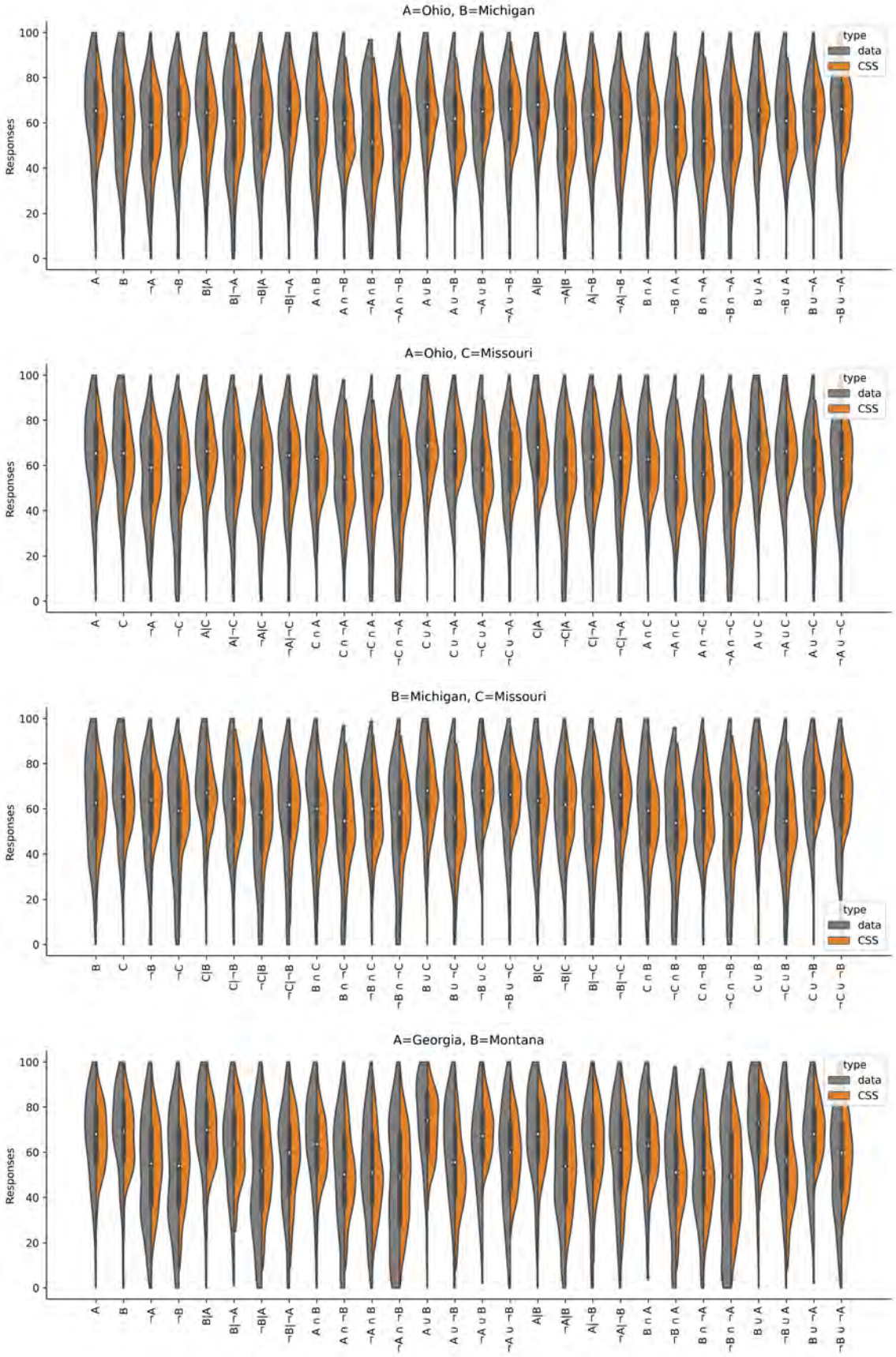
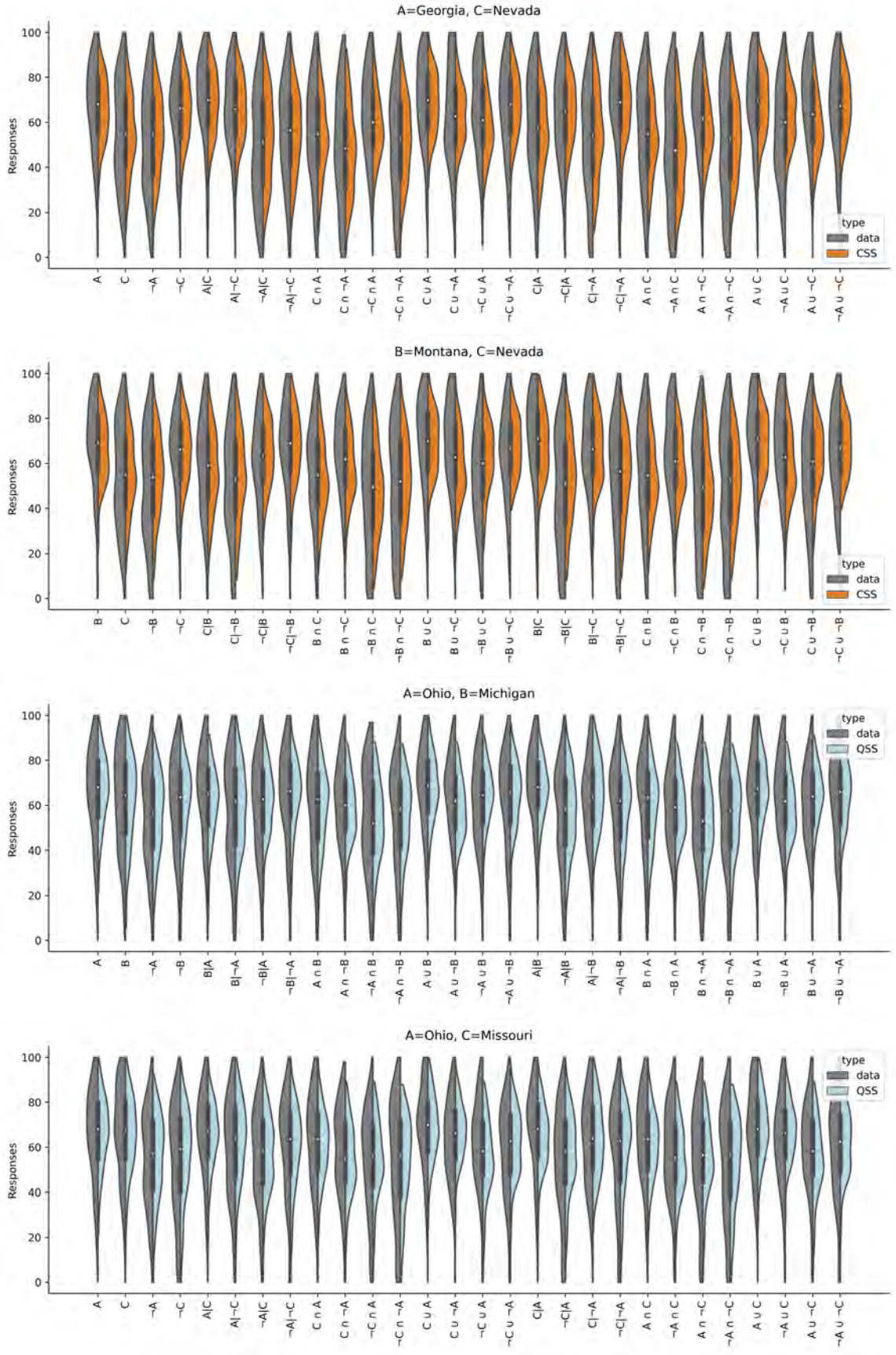
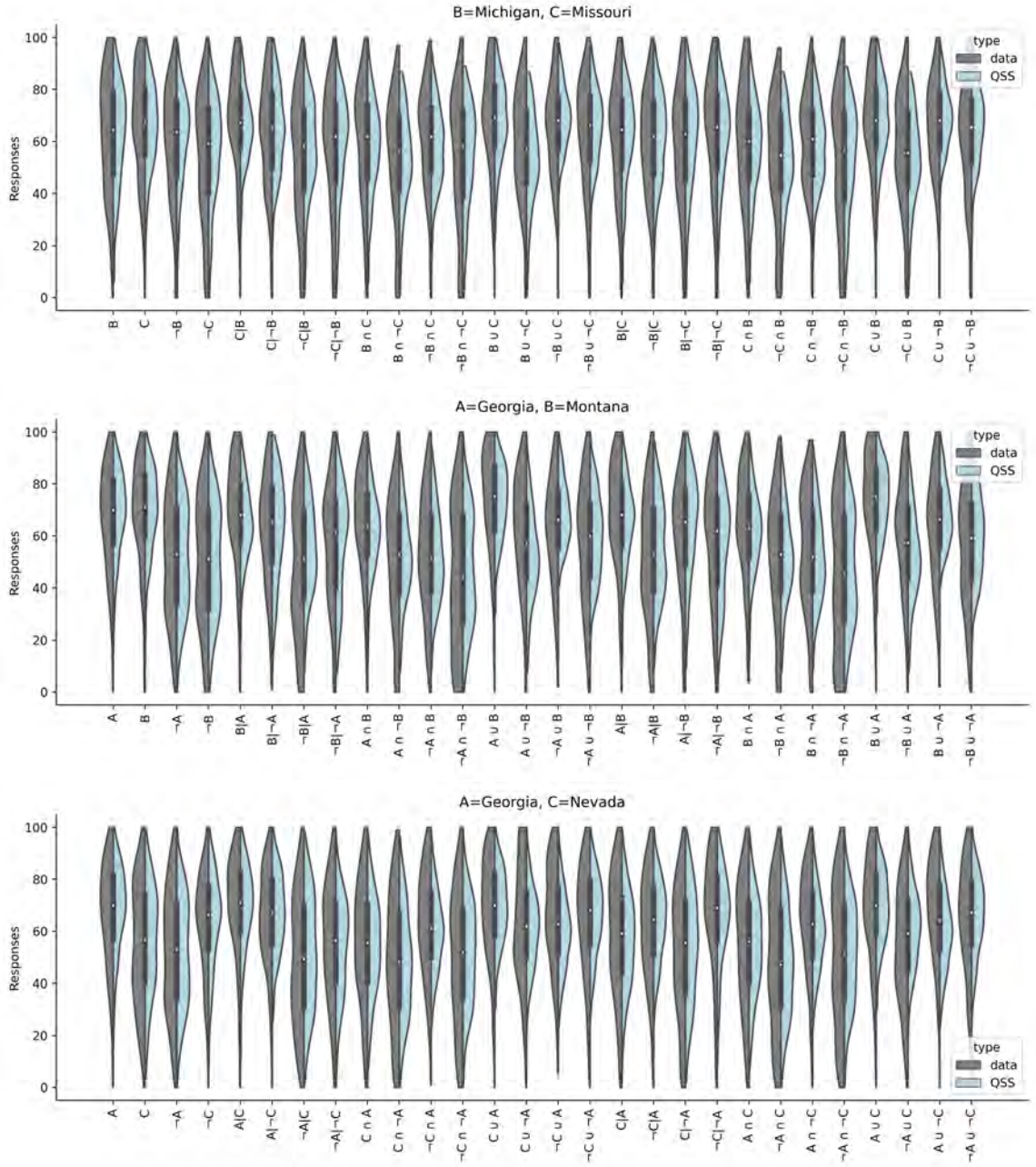


Figure S.5.2. Mean predictions of the quantum and classical variants of the Quantum Sequential Sampler against empirical results for participants whose quantum interference parameter is significant by the criterion ($p < .00008$).


A=Ohio, B=Michigan
A=Ohio, C=Missouri
B=Michigan, C=Missouri
A=Georgia, B=Montana
Responses
type
data
CSS


A=Georgia, C=Nevada
B=Montana, C=Nevada
A=Ohio, B=Michigan
A=Ohio, C=Missouri
Responses
type
data
CSS
QSS


B=Michigan, C=Missouri
A=Georgia, B=Montana
A=Georgia, C=Nevada
Responses
type
data
QSS

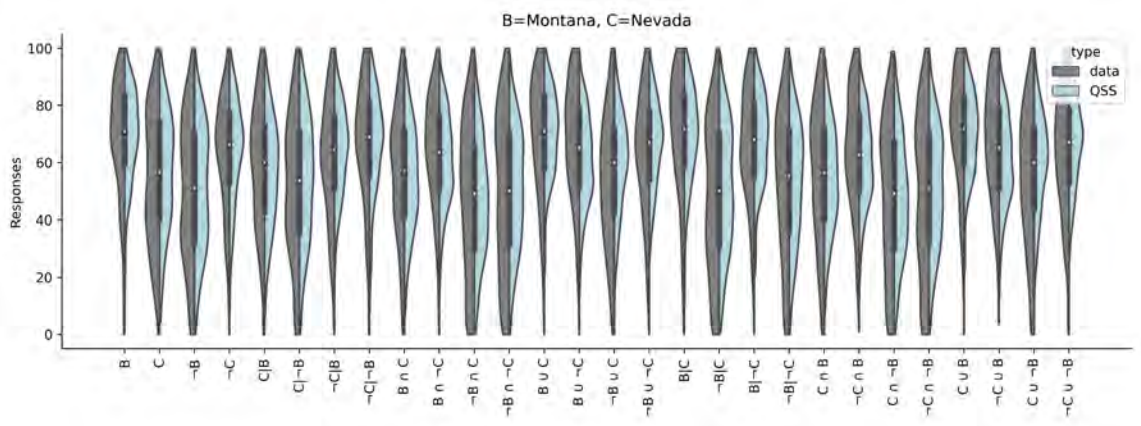
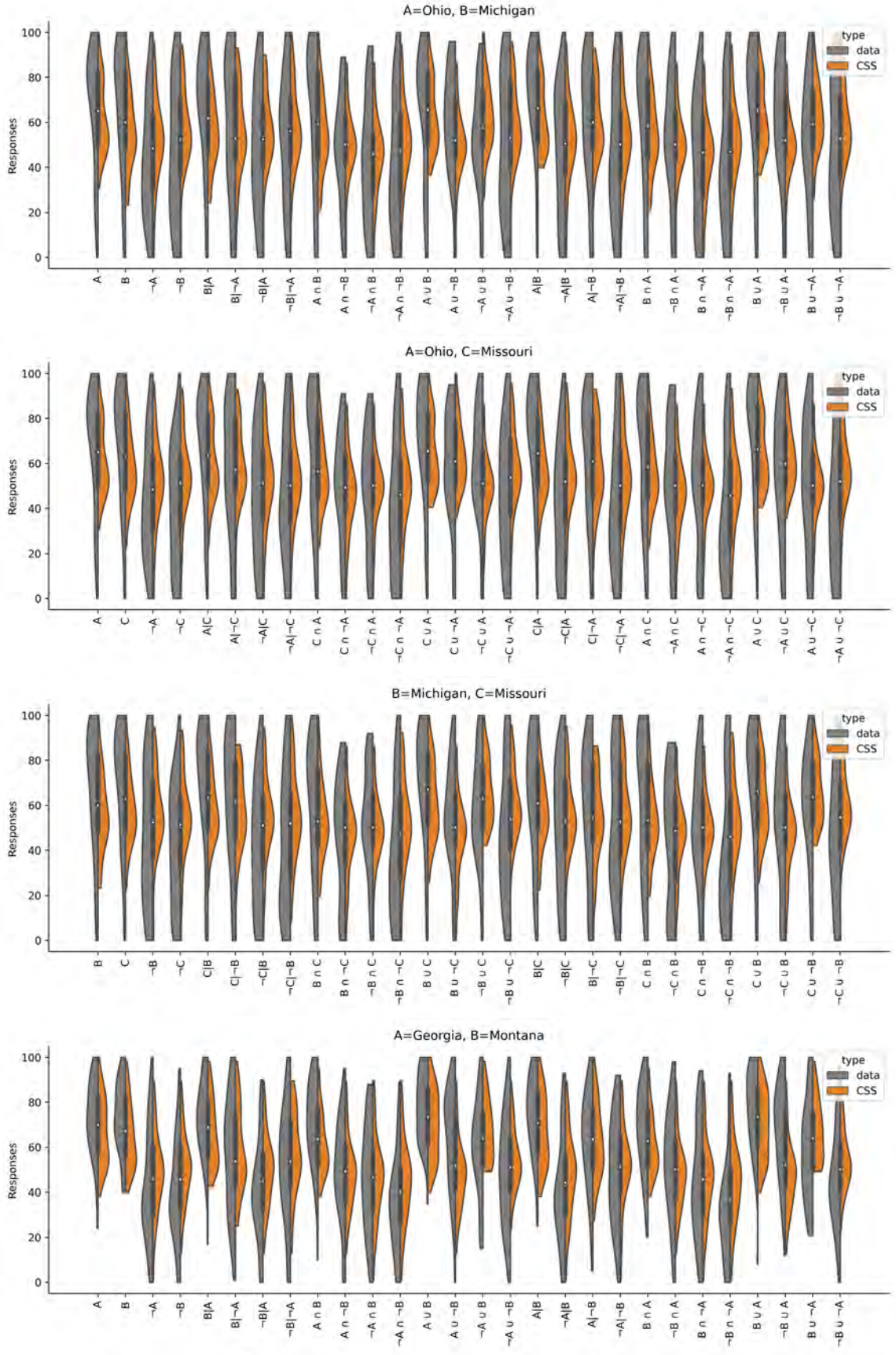
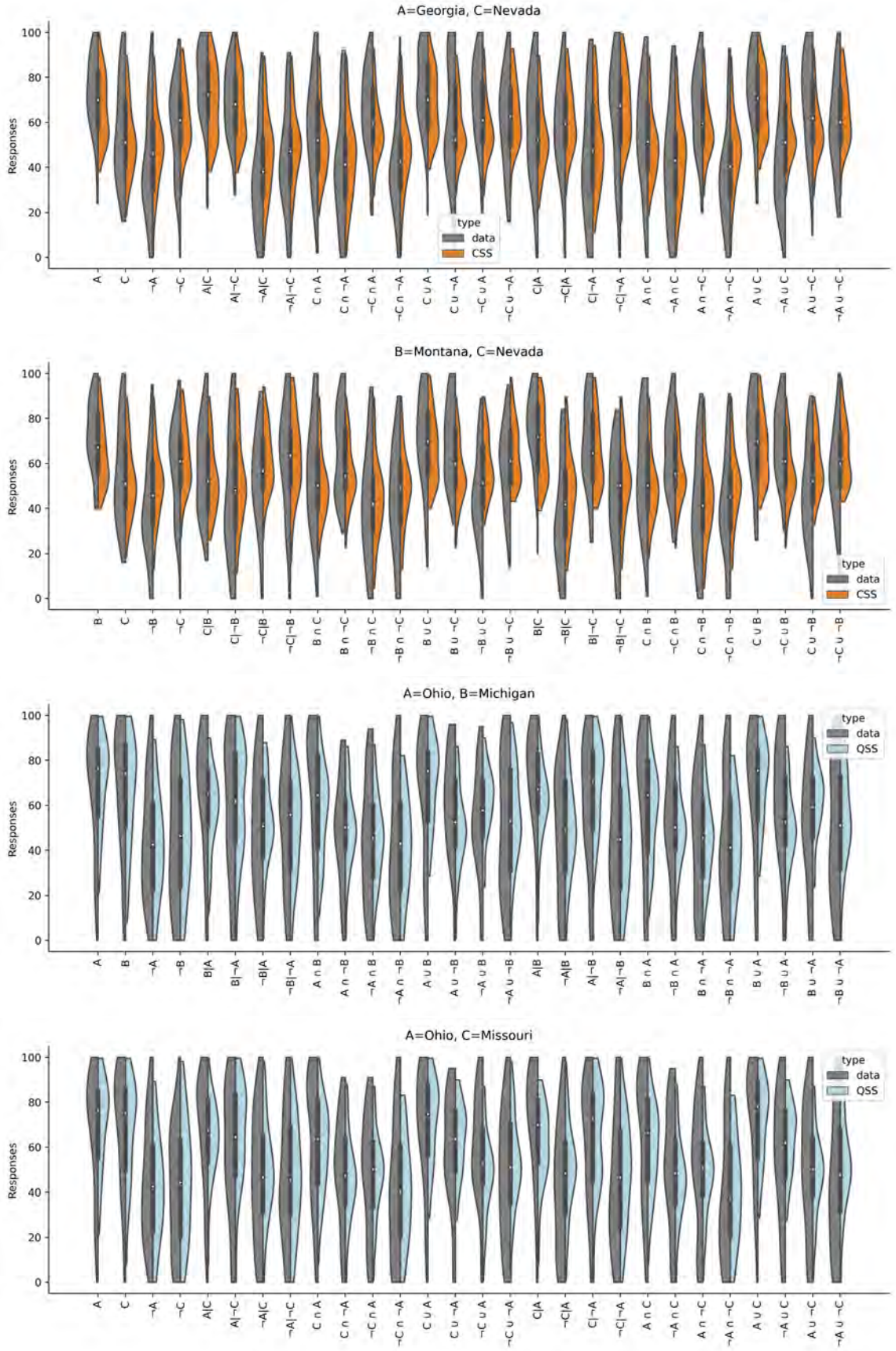
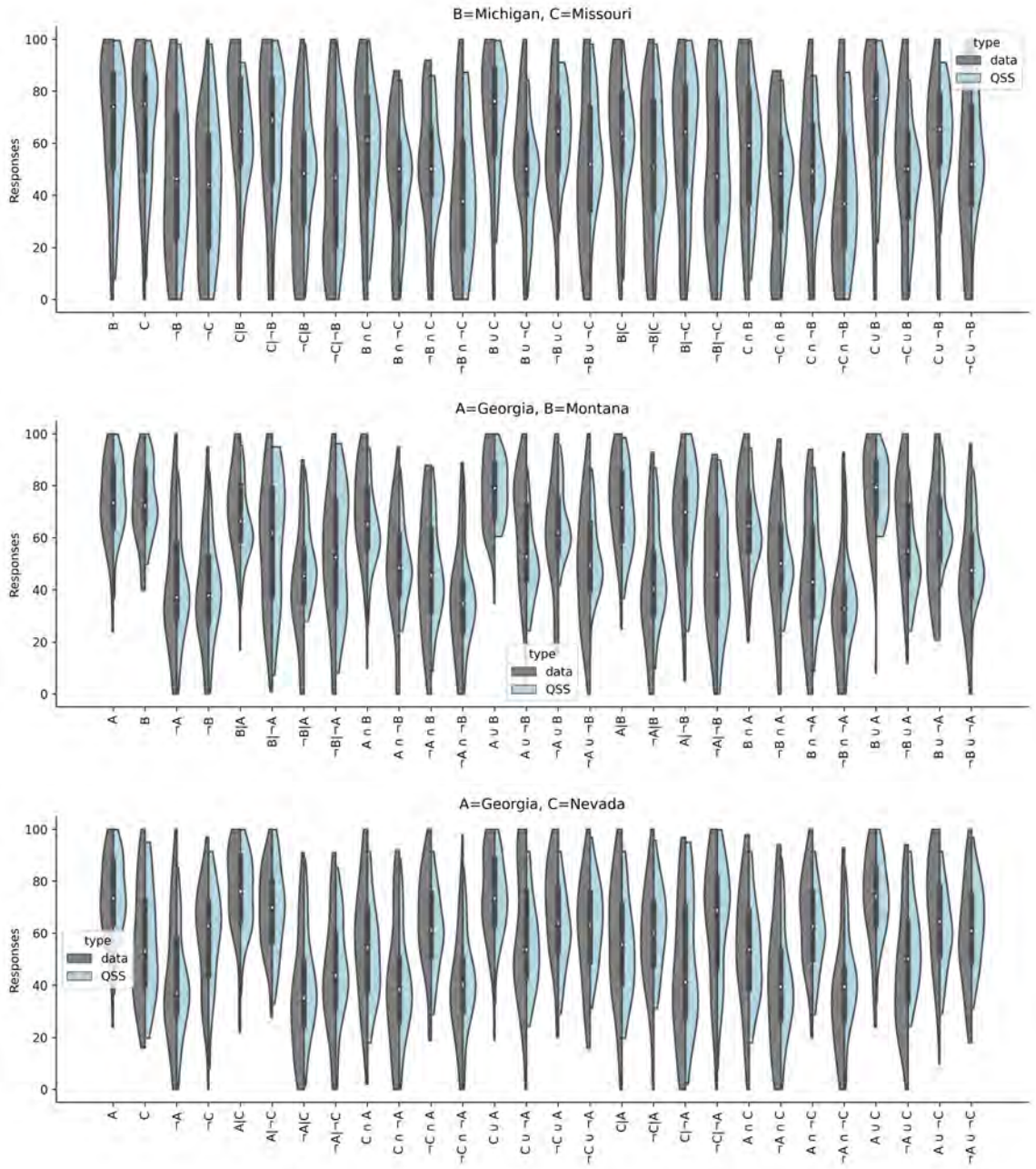


Figure S.5.3. Violin plots showing empirical data vs. the predictions of the classical and quantum variant of the Quantum Sequential Sampler model for for participants whose quantum interference parameter is significant by the criterion ($p < .05$).


A=Ohio, B=Michigan
A=Ohio, C=Missouri
B=Michigan, C=Missouri
A=Georgia, B=Montana
Responses
type
data
CSS


A=Georgia, C=Nevada
B=Montana, C=Nevada
A=Ohio, B=Michigan
A=Ohio, C=Missouri
Responses
type
data
CSS
QSS


B=Michigan, C=Missouri
A=Georgia, B=Montana
A=Georgia, C=Nevada
Responses
type
data
QSS

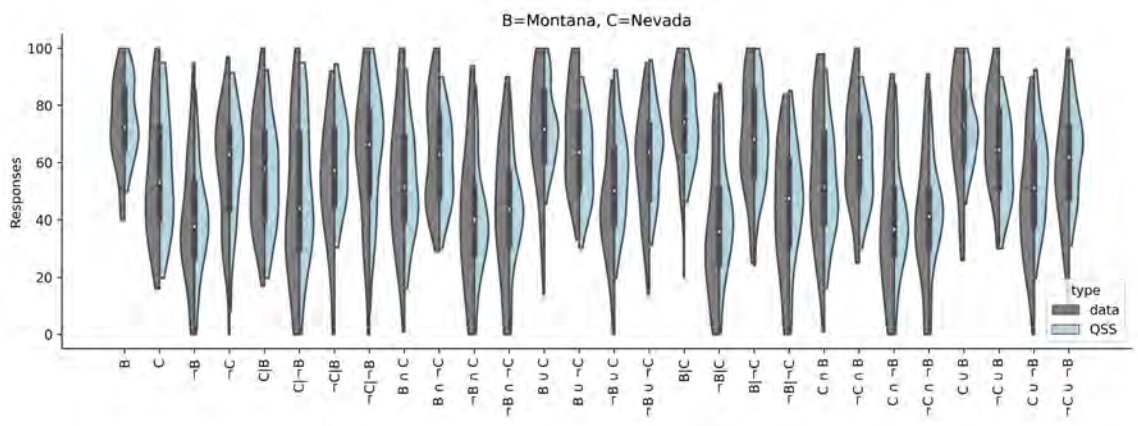


Figure S.5.4. Violin plots showing empirical data vs. the predictions of the classical and quantum variant of the Quantum Sequential Sampler model for for participants whose quantum interference parameter is significant by the criterion ($p < .00008$).

Supplementary Material 6

In this section, we showcase the kernel density plots for more participants whose quantum interference parameter is significant by ($p < .00008$).

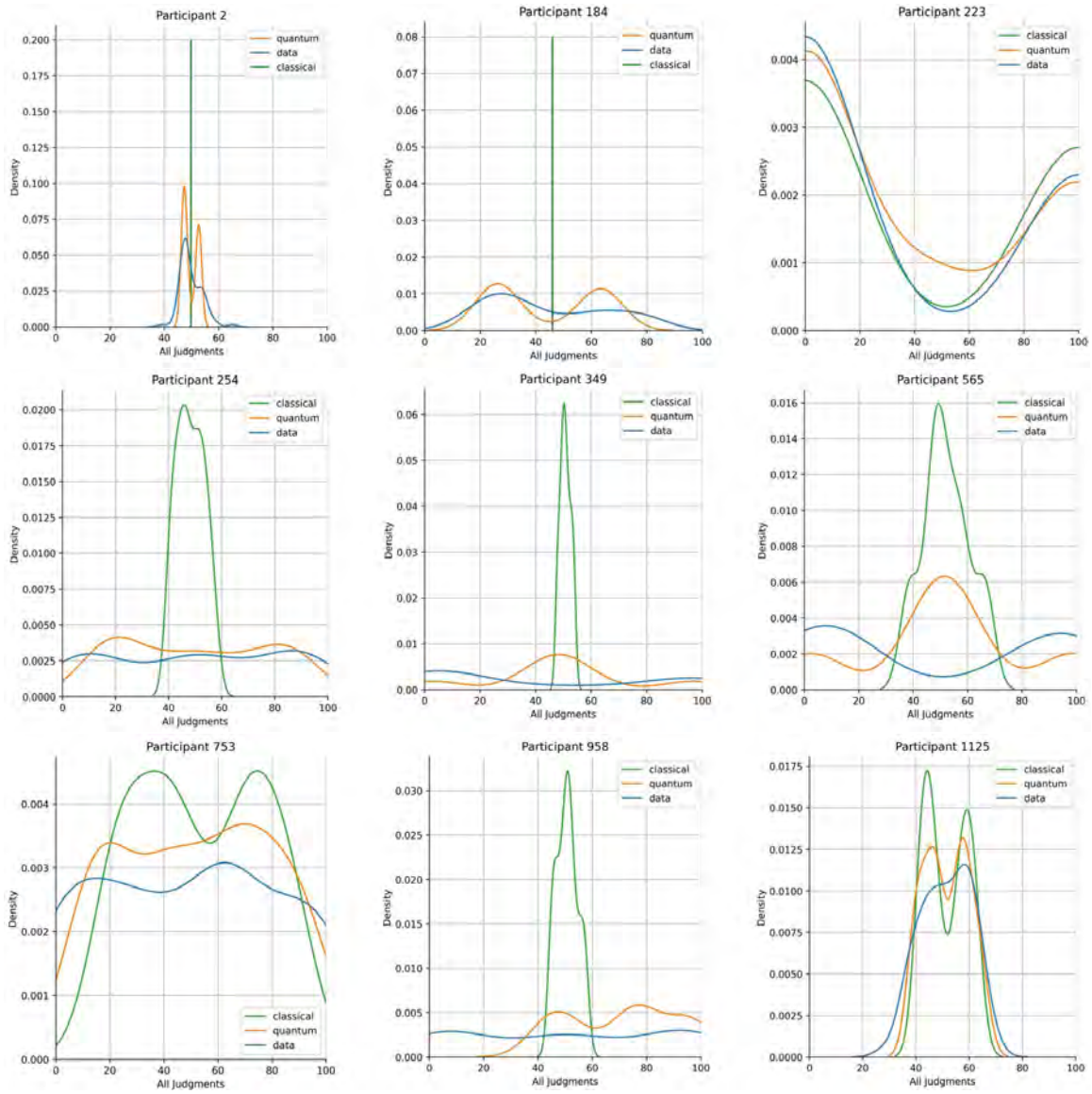


Figure S.6.1. Example of kernel denisty plots of individual whose quantum interference parameters are significant by the criterion ($p < .00008$). Classical refers to the classical variant of the quantum sequential sampler, and quantum refers to the quantum variant.

Supplementary Material 7

In the following analysis, the values for the Z identities Z_1 - Z_{18} were compared between the probabilistic judgments of election events collected to test the Quantum Sequential Sampler model, as well as the judgments of weather events collected by Zhu et al. (2020) to test their Bayesian Sampler model.

First, in the case of the Bayesian Sampler, participants provided three ratings for each question. We computed the average of these three ratings, so that there was a single score for each probability judgment. In two out of the five weather conditions in their experiment (normal/typical, icy/frosty, windy/cloudy, cold/rainy and warm/snowy) 59 participants provided probability ratings, while in the other three 84 participants provided ratings. This results in a total of 370 data points (probability judgments) available for the computation of the Z identities.

For the Quantum Sequential Sampler, data collection was more extensive and involved multiple events and all possible orders thereof. For that reason, each Z identity could be computed three times for each participant (six times if the order of events is considered relevant). To keep the comparison between the two data sets consistent, the various iterations of the same Z identity were averaged for each of the 1162 participants, before comparing against the Z identities computed on the basis of the Bayesian Sampler data.

The Z identities were compared using classical and Bayesian two-sided, independent samples t-tests. According to the Bayesian t-tests, there was very strong evidence that the values of the identities computed from the present data were different from those computed from the Zhu et al. (2020) data for all identities, except from Z_1, Z_9, Z_{16}, Z_{17}, as seen in Table S7.1.

The results using classical t-tests were identical (see Table S7.2). Please note that Levene's tests of homogeneity of variance indicated violations of the assumption of equal variances for most comparisons. As a consequence, Welch tests were also performed. There were no differences in the conclusions however and so, for simplicity, only the results using the classical t-tests are presented here.

The results are also shown in Figure S.7.1, highlighting the differences between the Z values computed from the two different data sets for most of the identities.

Bayesian Independent Samples T-Test	**BF_{10}**	**error %**
Z1	0.099	0.048
Z2	∞	0.000
Z3	3.586×10^{232}	2.565×10^{-236}
Z4	1.776×10^{25}	1.468×10^{-28}
Z5	7.227×10^{257}	8.398×10^{-263}
Z6	4.175×10^{236}	1.802×10^{-239}
Z7	7.268×10^{145}	2.204×10^{-151}
Z8	3.492×10^{119}	3.744×10^{-124}
Z9	0.100	0.047
Z10	4.820×10^{7}	7.740×10^{-11}
Z11	3.583×10^{24}	7.376×10^{-28}
Z12	1.837×10^{7}	2.051×10^{-10}
Z13	3.208×10^{27}	7.782×10^{-31}
Z14	2.875×10^{8}	1.275×10^{-11}
Z15	0.068	0.069
Z16	0.087	0.054
Z17	7.999×10^{6}	4.749×10^{-10}
Z18	2.601×10^{21}	1.080×10^{-24}

Table S7.1: Two-sided, independent samples Bayesian t-tests comparing the Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

Independent Samples T-Test	t	df	p
Z1	-0.887	1530	0.375
Z2	-51.981	1530	$< .001^a$
Z3	-39.630	1530	$< .001^a$
Z4	-11.312	1530	$< .001^a$
Z5	-42.618	1530	$< .001^a$
Z6	-40.109	1530	$< .001^a$
Z7	-29.277	1530	$< .001^a$
Z8	-25.978	1530	$< .001^a$
Z9	-0.906	1530	0.365
Z10	-6.468	1530	$< .001^a$
Z11	-11.156	1530	$< .001^a$
Z12	-6.311	1530	$< .001^a$
Z13	-11.804	1530	$< .001^a$
Z14	6.749	1530	$< .001^a$
Z15	-0.179	1530	0.858
Z16	-0.724	1530	0.469
Z17	-6.173	1530	$< .001^a$
Z18	-10.430	1530	$< .001^a$

Table S7.2: Two-sided, independent samples classical t-tests comparing the Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

[a] Levene's test is significant ($p < .05$), suggesting a violation of the equal variances assumption

Identity	**Mean**$_{BS}$	**Mean**$_{ass}$	**SD**$_{BS}$	**SD**$_{QSS}$
Z1	-0.025	-0.014	0.195	0.212
Z2	-0.371	0.000	0.215	0.066
Z3	-0.028	0.508	0.124	0.250
Z4	0.344	0.508	0.216	0.251
Z5	-0.084	0.522	0.240	0.238
Z6	-0.002	0.523	0.154	0.236
Z7	0.342	1.031	0.224	0.435
Z8	0.367	1.045	0.314	0.470
Z9	-0.005	-0.001	0.092	0.069
Z10	0.067	0.153	0.250	0.214
Z11	0.019	0.152	0.148	0.215
Z12	0.072	0.154	0.236	0.212
Z13	0.014	0.151	0.124	0.213
Z14	0.054	0.002	0.231	0.067
Z15	0.162	0.163	0.142	0.107
Z16	0.157	0.162	0.142	0.116
Z17	0.229	0.317	0.229	0.240
Z18	0.176	0.315	0.172	0.237

Table S7.3: Descriptive data for the Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

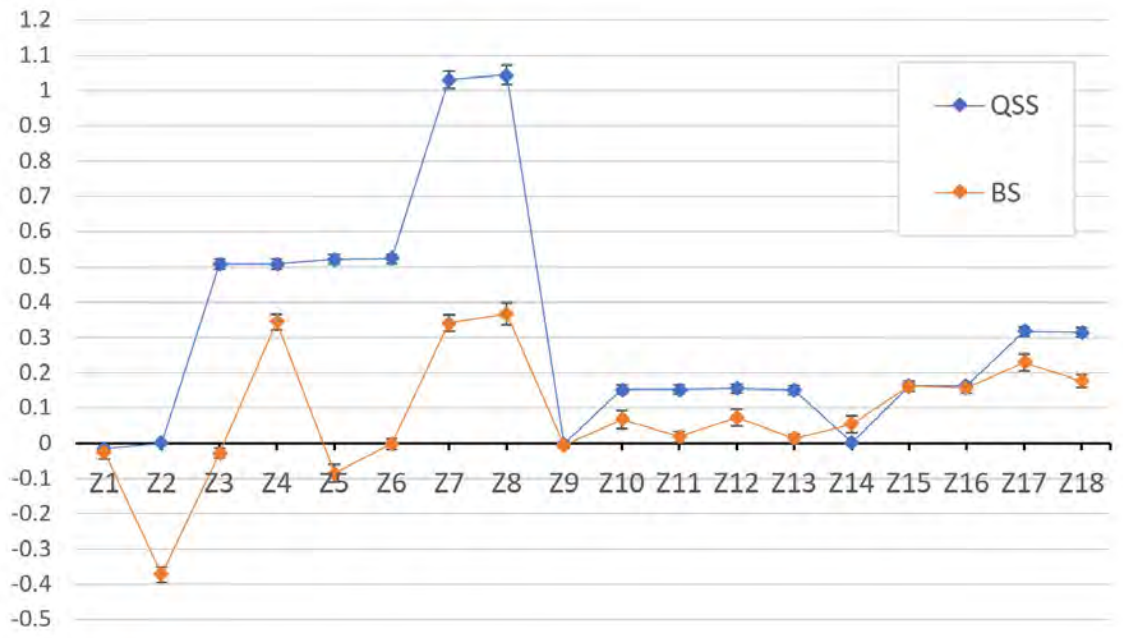
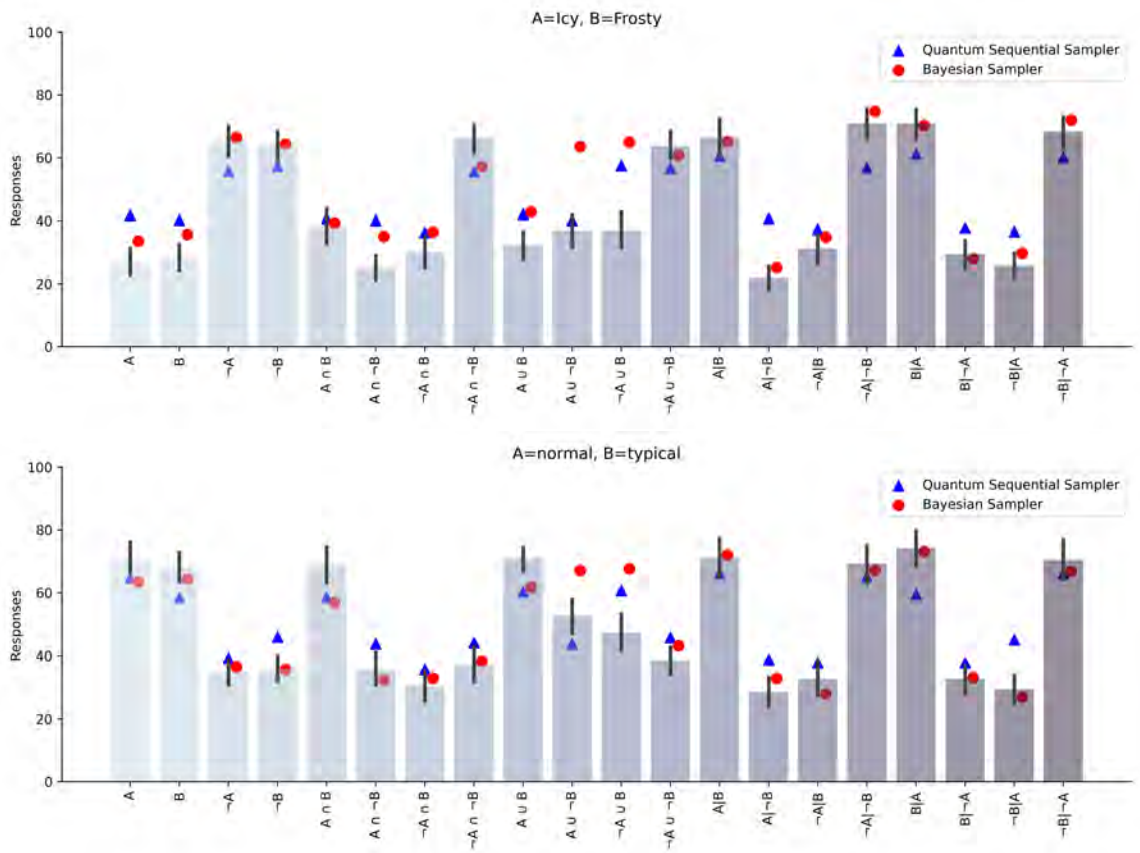


Figure S.7.1. Z identities computed from the data for the Quantum Sequential Sampler vs the Bayesian Sampler.

Supplementary Material 8

In this section, we present all analysis plots that compare the Quantum Sequential Sampler (quantum variant) with the Bayesian Sampler model, using the dataset from Zhu et al. (2020) which includes repeated probability judgment measurements. It should be noted, as detailed in Appendix 2, that we have applied a rounding mechanism to the nearest 5 or 10 during our model fitting process to account for the significant presence of this rounding bias in Zhu et al.'s dataset. Consequently, the Bayesian Sampler's predictions in our analysis may slightly deviate from those reported in Zhu et al. (2020). Nonetheless, upon examination, it is evident that these differences are minor to be seen in plots and do not alter any analytical trend.



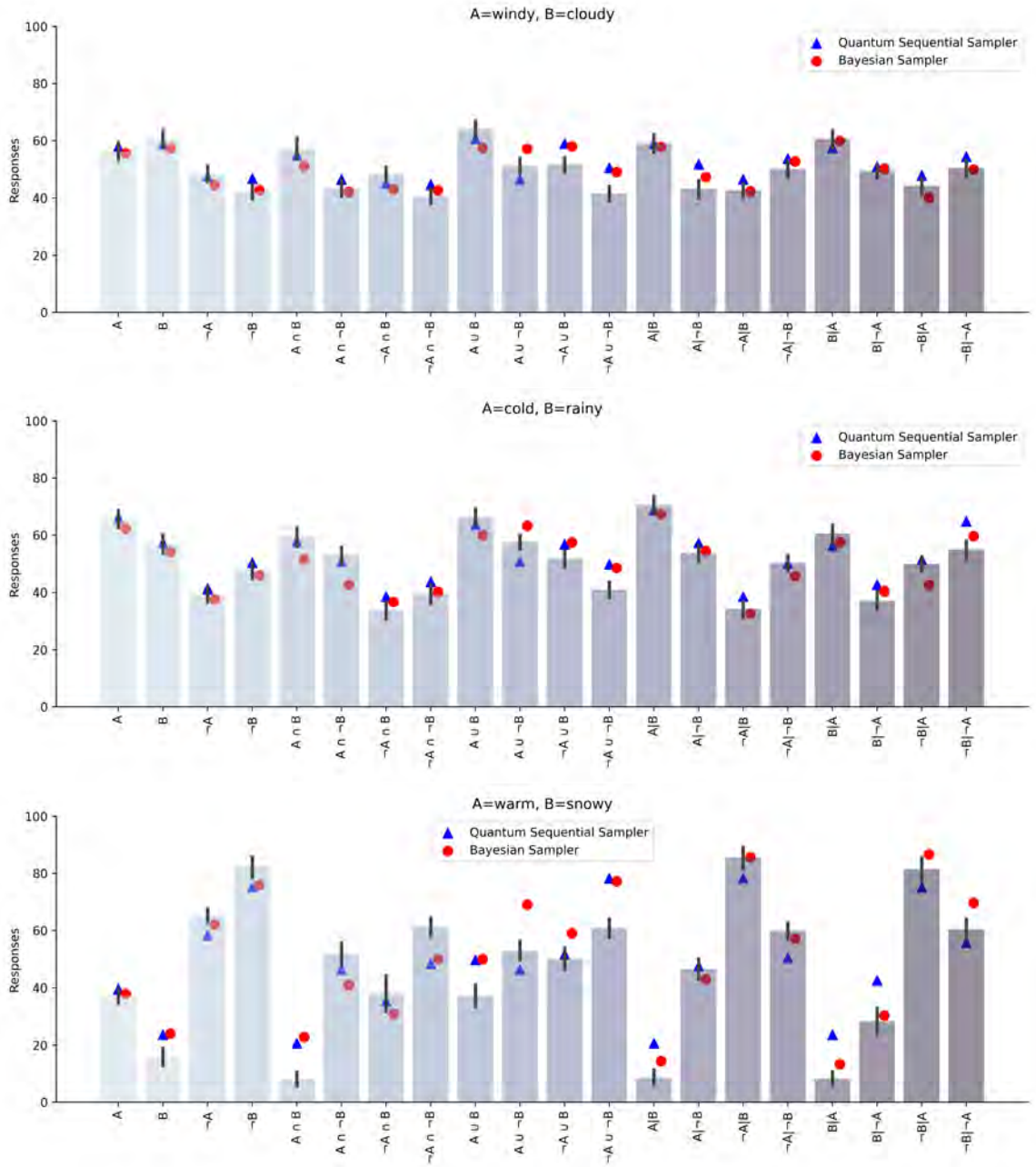
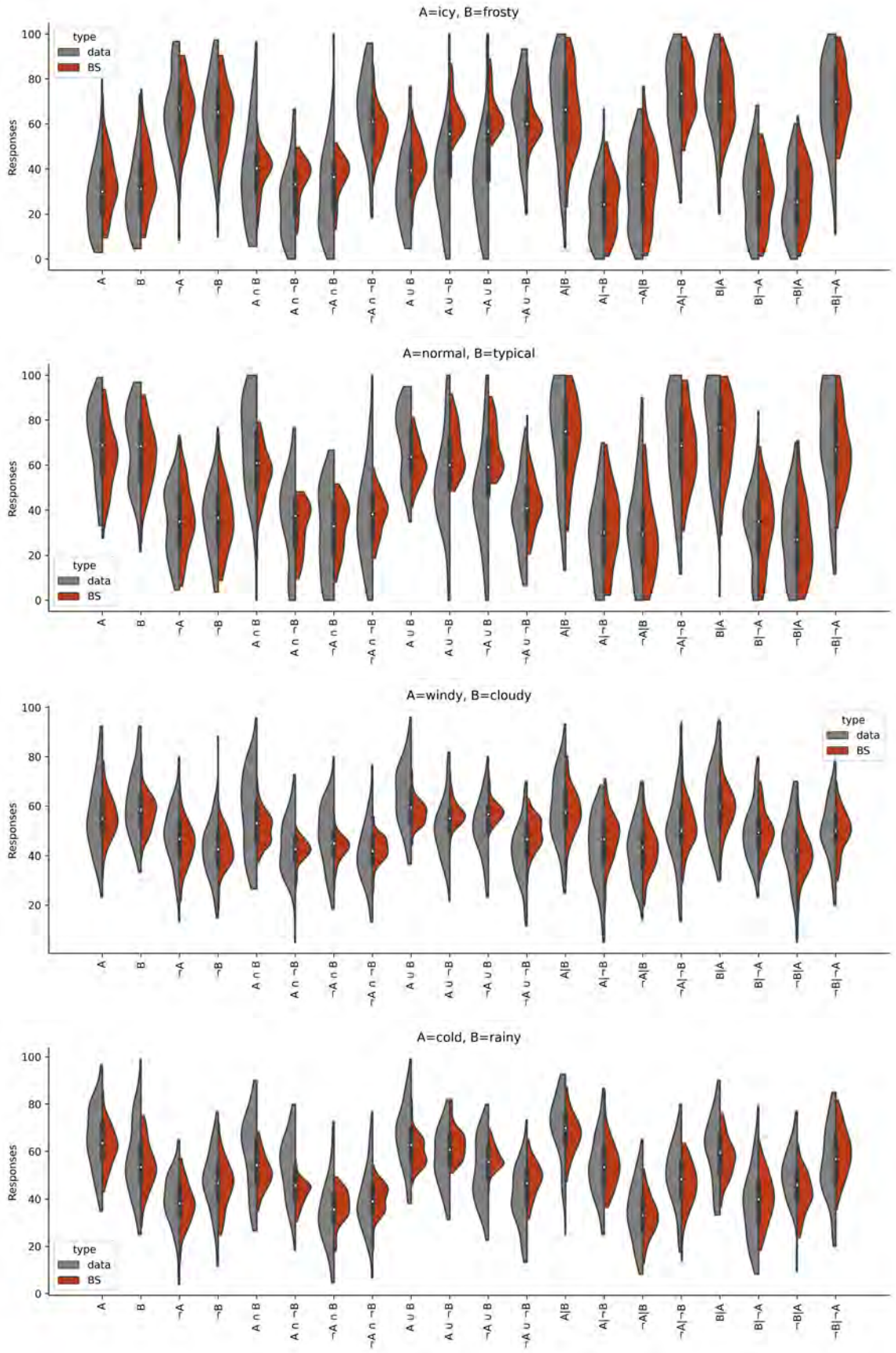


Figure S.8.1. Mean predictions of the Bayesian Sampler and the Quantum Sequential Sampler against empirical results in Zhu et al. (2020).


A=icy, B=frosty
A=normal, B=typical
A=windy, B=cloudy
A=cold, B=rainy
type
data
BS
Responses
A
B
¬A
¬B
A ∩ B
A ∩ ¬B
¬A ∩ B
¬A ∩ ¬B
A ∪ B
A ∪ ¬B
¬A ∪ B
¬A ∪ ¬B
A|B
A|¬B
¬A|B
¬A|¬B
B|A
B|¬A
¬B|A
¬B|¬A

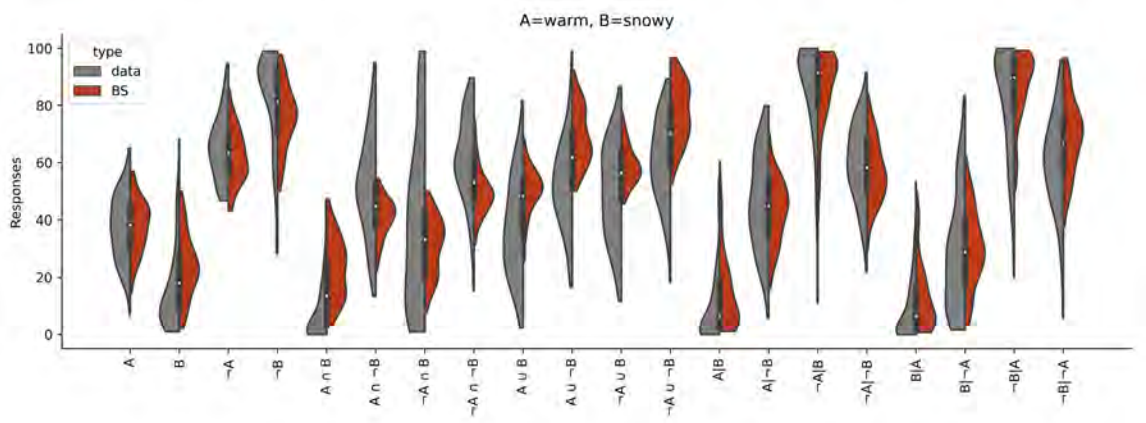
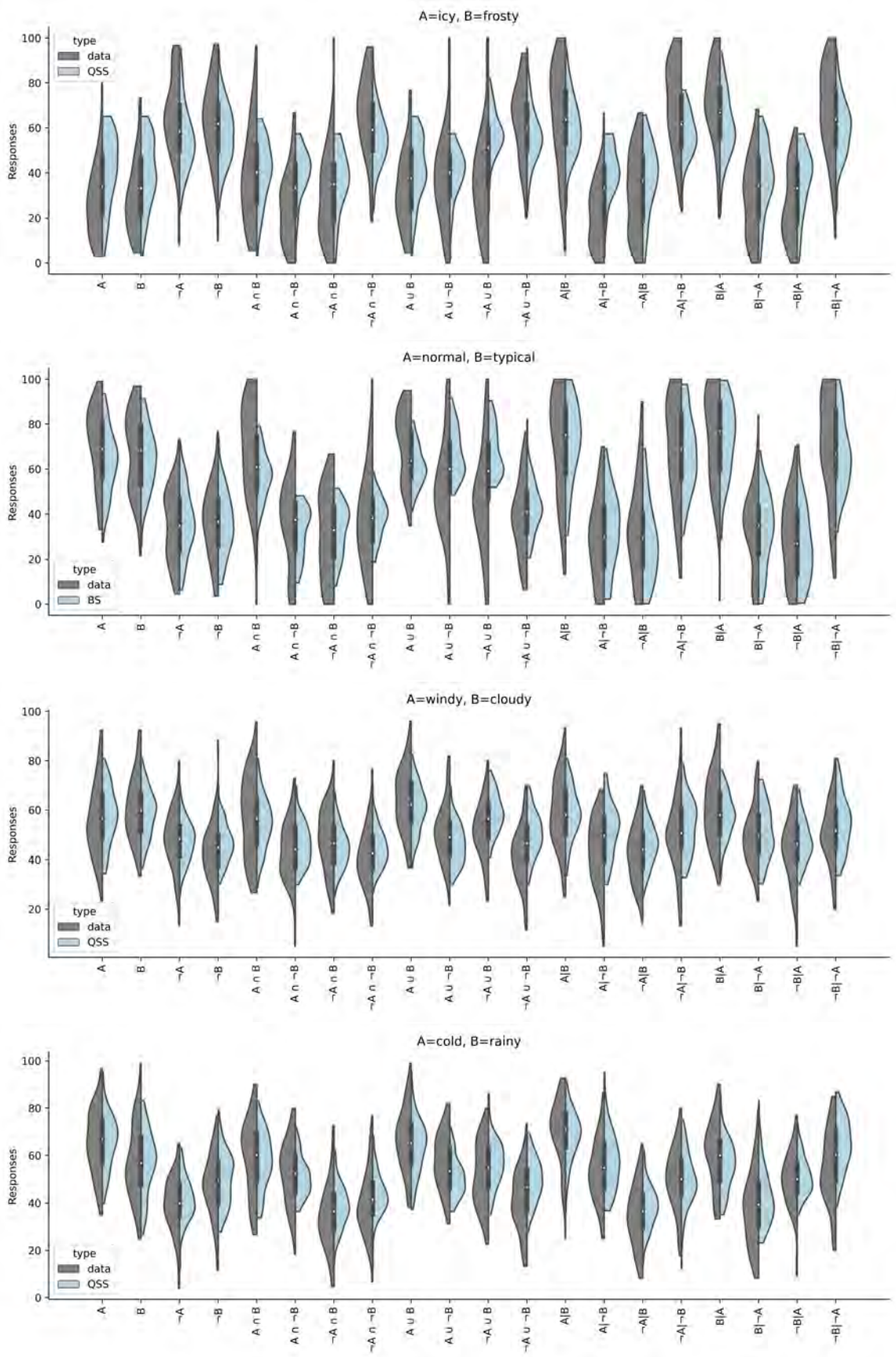


Figure S.8.2. Violin plots showing empirical data of Zhu et al. (2020) vs. the predictions of the Bayesian Sampler.


A=icy, B=frosty
A=normal, B=typical
A=windy, B=cloudy
A=cold, B=rainy
Responses
type
data
QSS
BS
A
B
¬A
¬B
A ∩ B
A ∩ ¬B
¬A ∩ B
¬A ∩ ¬B
A ∪ B
A ∪ ¬B
¬A ∪ B
¬A ∪ ¬B
A|B
A|¬B
¬A|B
¬A|¬B
B|A
B|¬A
¬B|A
¬B|¬A

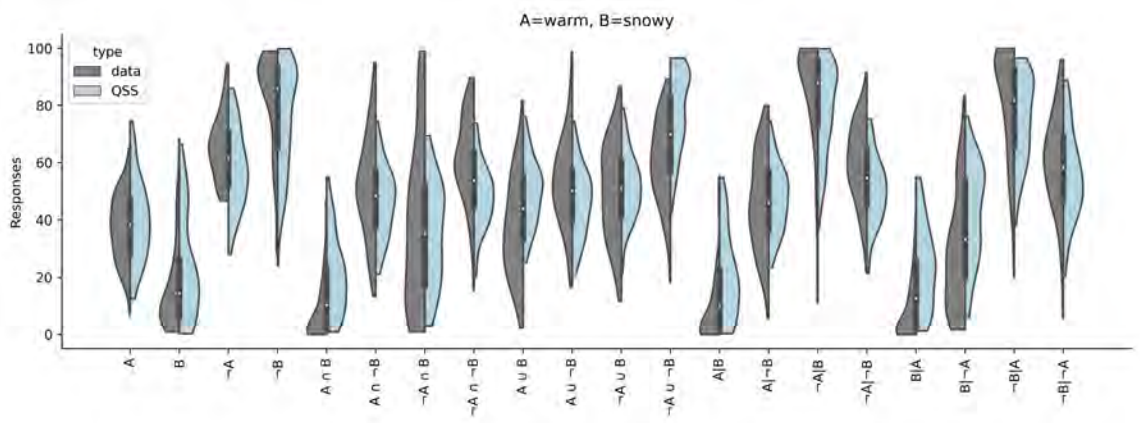
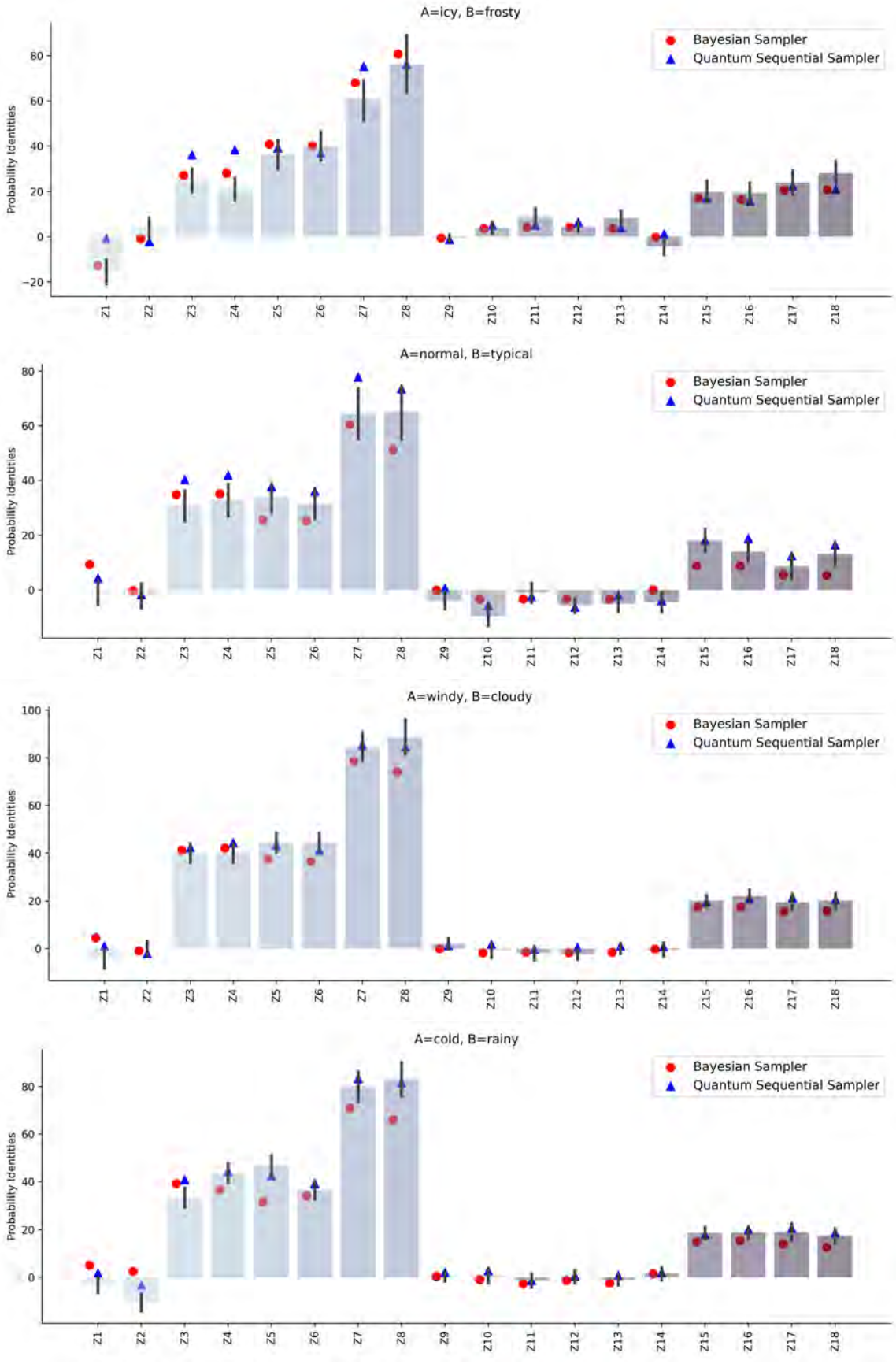


Figure S.8.3. Violin plots showing empirical data of Zhu et al. (2020) vs. the predictions of the Quantum Sequential Sampler.


A=icy, B=frosty
A=normal, B=typical
A=windy, B=cloudy
A=cold, B=rainy
Bayesian Sampler
Quantum Sequential Sampler
Probability Identities
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Z9
Z10
Z11
Z12
Z13
Z14
Z15
Z16
Z17
Z18

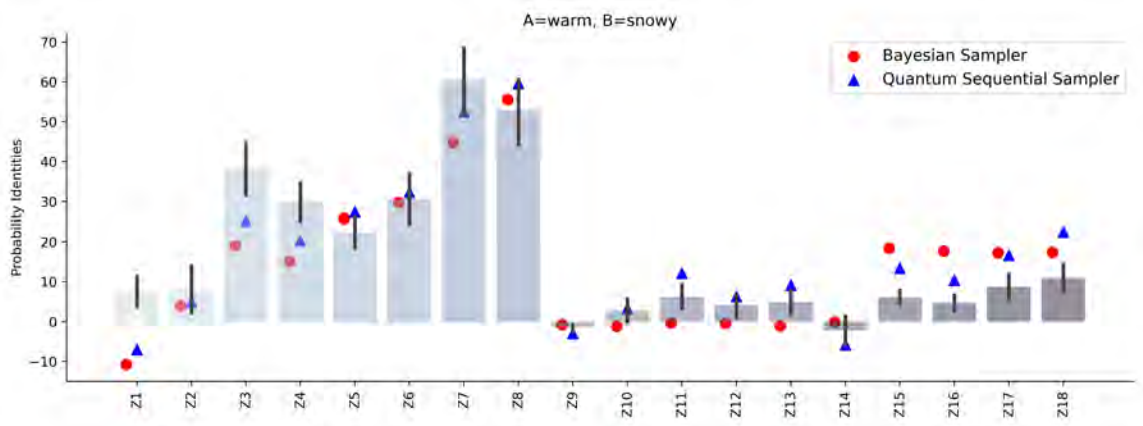


Figure S.8.4. The empirical value of probabilistic identities in Zhu et al. (2020), together with values computed from best-fit predictions, from the Bayesian Sampler and the Quantum Sequential Sampler.