Context effects produced by question orders reveal quantum nature of human judgments

Zheng Wang, Tyler Solloway, Richard M. Shiffrin, and Jerome R. Busemeyer

The hypothesis that human reasoning obeys the laws of quantum rather than classical probability has been used in recent years to explain a variety of seemingly “irrational” judgment and decision-making findings. This article provides independent evidence for this hypothesis based on an a priori prediction, called the quantum question (QQ) equality, concerning the effect of asking attitude questions successively in different orders. We empirically evaluated the predicted QQ equality using 70 national representative surveys and two laboratory experiments that manipulated question orders. Each national study contained 651–3,006 participants. The results provided strong support for the predicted QQ equality. These findings suggest that quantum probability theory, initially invented to explain noncommutativity of measurements in physics, provides a simple account for a surprising regularity regarding measurement order effects in social and behavioral science.

Understanding human reasoning under uncertainty is fundamental for improving decisions about environmental policies, economic planning, public health, and many other important areas. Fifty years of behavioral decision-making research has established that humans do not always follow the “rational” rules of Bayesian probability theory (1). Recently, a group of psychologists and physicists have formulated new rules for human reasoning under uncertainty based on quantum probability theory (2–4). This article reports a test of this theory based on results from a quite different paradigm. We show that the theory implies an a priori and precise prediction called the quantum question (QQ) equality (5). This parameter-free prediction concerns the effect of question order on human judgments. The prediction was tested with the results of 70 national representative surveys, most containing more than 1,000 participants per survey, and two laboratory studies, that manipulated question order. This article presents the QQ equality, its surprisingly strong empirical support, and the key quantum principle, called the law of reciprocity, upon which the QQ equality was mathematically derived. Finally, we explain why human judgments follow quantum rules even if the brain may not be a quantum computer.

The QQ Equality

To introduce the QQ equality, consider three examples of context effects on answers to attitude questions in surveys, illustrated in Table 1. These are the results of three Gallup polls reported in a seminal article on question order effects (6). Each poll included a representative sample of around 1,000 US adults. The participants in one random half of the sample were asked two questions in one order, and those in the other half were asked the same two questions in the opposite order. In the first poll, people were asked whether Bill Clinton was honest and trustworthy, and whether Al Gore was honest and trustworthy. In the second poll, people were asked whether or not Pete Rose should be admitted to the baseball hall of fame, and whether or not shoeless Joe Jackson should be admitted to the baseball hall of fame. Each column of three two-way tables presents the results from one of the three polls. The cells within the top two tables in each column show the observed proportions for the four response combinations for each question order. For each poll, a “context effect” produced by the question order occurs when any of the four proportions in the top table differs from its corresponding cell in the middle table; these differences are shown in the bottom table. A rigorous statistical test of the four context effects for a single poll can be measured using a χ² statistic (SI Text). As shown by the χ² statistic on the order effects, all three polls produced large and statistically significant (p < 0.05) order effects, but with strikingly different patterns. [Note that these order effects go against the commutative property of joint probability: P(Yes A ∩ Yes B) = P(Yes A) · P(Yes B|Yes A) = P(Yes B) · P(Yes A|Yes B) = P(Yes B|Yes A).]

Despite the different patterns of context effects displayed in the first two polls in Table 1, they both reveal an interesting common property: The sum of the context effects across the two cells within a common diagonal is close to zero. We call this sum the q value. (It is a mathematical property of any context effect table that the q value produced by the main diagonal is always equal but opposite in sign to the q value produced by the off-diagonal; SI Text). Our quantum theory, presented later, predicts that the expectation of the q value equals zero for these two polls, E(q) = 0, which we call the “QQ equality.” For example, for the first poll, the q value computed by summing the context effects for the two off-diagonal cells equals −0.003; the corresponding q value for the second poll equals −0.020. Psychologically, this

Significance

In recent years, quantum probability theory has been used to explain a range of seemingly irrational human decision-making behaviors. The quantum models generally outperform traditional models in fitting human data, but both modeling approaches require optimizing parameter values. However, quantum theory makes a universal, nonparametric prediction for differing outcomes when two successive questions (e.g., attitude judgments) are asked in different orders. Quite remarkably, this prediction was strongly upheld in 70 national surveys carried out over the last decade (and in two laboratory experiments) and is not one derivable by any known cognitive constraints. The findings lend strong support to the idea that human decision making may be based on quantum probability.
means that the number of people who switch from “yes–yes” to “no–no” must be offset by the number who switch in the opposite direction; likewise, the number of people who switch from “yes–no” to “no–yes” must be offset by the number who switch in the opposite direction. (The QQ equality is predicted to hold no matter whether context/order effects occur on neither, one, or both diagonals.) To our knowledge, no traditional psychological theories impose this precise kind of symmetry constraint on context effects.

There is no mathematical constraint that forces the $q$ value to be zero. [In general, the set of all context effects forms a three-dimensional pyramid, but those that satisfy the QQ equality forms only a triangular plane that lies within the pyramid (SI Text).] For example, the third poll was selected to be presented in Table 1 because our quantum model predicts that it should produce a violation of the QQ equality, and indeed, the $q$ value is big and equals $–0.15$. As discussed later, a key assumption required for the derivation of the QQ equality is that the questions being examined are asked successively with no additional information inserted before or between questions. In our examples in Table 1, the first two polls satisfied this condition, but the third did not. A rigorous statistical test of the QQ equality can be computed using a $\chi^2$ test, which we call the $q$ test (SI Text). As shown in Table 1, as predicted by our quantum theory, the $q$ tests for the first two polls are not statistically different from zero, but that for the third poll is.

To assess the generality of the QQ equality, we obtained a total of 70 national representative surveys that manipulated the order of questions when they were asked successively. Most of them included more than 1,000 respondents. In addition, we included two laboratory experiments by the authors, each including more than 100 respondents (SI Text). Of the 70 national surveys, two of them were the polls shown in Table 1 (the third poll in Table 1 was excluded for the reason mentioned above); a third was another Gallop poll reported by Moore (6); and a fourth was a classic study on context effects (7) that provided the complete two-way tables required for the $q$ test. (Unfortunately, most studies on context effects only report the marginal proportions rather than the complete two-way table required for conducting the $q$ test. The remaining 66 national surveys were all of the available studies conducted by Pew Research Center on various topics during the decade of 2001–2011 that manipulated the order of two questions. (See SI Text for more details about the 72 studies.)

The surprising nature of the results concerning the QQ equality can be illustrated in several ways. Because the QQ equality predicts that the sum of the two diagonal entries is zero, it is informative to plot one entry against the other. For each study, we selected the diagonal producing the larger “order effect” (defined as the sum of the absolute values of the two diagonal entries). For example, as seen in Table 1, the larger order effect for the Clinton–Gore poll occurs on the main diagonal, which produced the pair of context effects $x = -0.0726$ and $y = 0.0192$, and the size of the order effect for this pair equals 0.1482. For the white–black poll, the larger order effect occurs on the minor diagonal, which produced the pair of context effects $x = -0.1205$ and $y = 0.1015$, and the size of the order effect for this pair equals 0.2220. Each of the 72 points in Fig. 1, Left plots these two $(x, y)$ values from a dataset: The horizontal axis represents the $x$ context effect, and the vertical axis represents the $y$ context effect. More extreme order effects produce more extreme values on the $x, y$ axes. If we ignore the QQ equality, there should not be any a priori constraints on the relations between the pairs, and hence, there is no reason to expect any particular correlation between them. However, according to the QQ equality, the context effect in one cell of a diagonal should be exactly the negative of the context effect in the other cell within the same diagonal, and thus all these pairs should fall along a line with the intercept of 0 and the slope of $–1$, producing a perfect negative correlation. Fig. 1, Left shows the scatter plot of the 72 pairs of context effects from all 72 studies. The straight line in the figure is not a fitted regression line; instead, it is the a priori predicted line with the intercept of 0 and the slope of $–1$. As shown, the data points fall closely along this predicted line, and the correlation $r = –0.82 (r = –0.73$, when two extreme values are excluded). The surprising regularity illustrated in this scatter plot provides strong support for the QQ equality prediction.

Another question concerns the possible range of $q$ values. The finding that the $q$ value remains close to zero is interesting only if the range of its possible values is much larger than the observed $q$ values. The third example of Rose–Jackson poll in Table 1 demonstrates that a large $q$ value can occur, but it is important to

Table 1. $\chi^2$ results for three Gallup survey experiments reported in a seminal article on question order effects (6)

<table>
<thead>
<tr>
<th></th>
<th>Clinton–Gore</th>
<th>White–black</th>
<th>Rose–Jackson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gy</td>
<td>By</td>
<td>Gy</td>
</tr>
<tr>
<td></td>
<td>Gn</td>
<td>Bn</td>
<td>Gn</td>
</tr>
<tr>
<td>Cy</td>
<td>0.4899</td>
<td>0.0447</td>
<td>0.2886</td>
</tr>
<tr>
<td>Cy</td>
<td>0.5625</td>
<td>0.0255</td>
<td>0.4012</td>
</tr>
<tr>
<td>Cy</td>
<td>–0.0726</td>
<td>0.0192</td>
<td>–0.0025</td>
</tr>
<tr>
<td>Cn</td>
<td>0.1991</td>
<td>0.2130</td>
<td>0.0528</td>
</tr>
<tr>
<td>Cn</td>
<td>–0.0224</td>
<td>0.0756</td>
<td>0.0777</td>
</tr>
</tbody>
</table>

Each column presents the results from one survey. In each column, the top two-way table shows the observed proportions from one question order, the middle two-way table shows those from the other question order, and the bottom two-way table summarizes the context effects. Context effects were computed by subtracting the observed response proportion in each cell obtained in the AB (e.g., Clinton–Gore) order by that in the BA (e.g., Gore–Clinton) order.
examine this issue for all 72 studies. For any observed table of context effects, we can bound the q value by the size of the order effect. Recall that we defined the size of the order effect in terms of the diagonal with the larger summed absolute values of context effects (e.g., 0.15 for Clinton–Gore and 0.22 for white-black). The q value can possibly equal but cannot exceed the size of the order effect defined in this manner (SI Text). If the order effect is close to zero, then the q value must also be close to zero, and sampling estimation error for both will cause them to be approximately equal in size, so the q test is only interesting when the order effects are well above zero. The relation between the q value and the order effect can be described by their ratio: size of the q value/size of the order effect. Because of the sampling error, this ratio will necessarily be close to 1 when the order effect is very small, but if the QQ equality holds, then this ratio should drop to zero as the size of the order effect increases. Fig. 1, Right plots this ratio for the 17 studies that produced an order effect greater than 0.10. As predicted by the QQ equality and shown in the figure, this ratio starts well below 1.0 when the order effect is small and drops toward zero when the order effect becomes large; over the entire range, the q value remains small.

Fig. 1 may provide a compelling illustration, but it does not substitute for an appropriate statistical test of the null hypothesis that the expectation of the q values is zero. We exclude the four national surveys that were specifically selected from previous studies because they found question order effects (although including them does not change our conclusions below; SI Text); and we analyze the distributions of χ² statistics for order effects and q values from the remaining 66 Pew surveys that were selected without any bias for either test. As described before, these include all of the datasets available from Pew in the past decade that manipulated the order of two questions. On the one hand, the χ² distribution test for order effects produced a significant deviation from the null hypothesis (p = 0.0004); on the other hand, the χ² distribution test for the q values indicates no significant deviation from the null hypothesis (p = 0.4625) (see SI Text for detail on the χ² tests). Taken together, these results show that across all 66 Pew datasets, there are significant question order effects, and the QQ equality holds as predicted.

In summary, we have presented strong evidence that context effects produced by the order of questions satisfy the QQ equality predicted by quantum theory: (i) The context effect from one cell of a diagonal is negatively correlated with that from the other cell; (ii) the q value remains small even as the size of the context/order effect increases; (iii) the q values do not differ significantly from zero as tested by a large set of national survey data on various topics collected in the past decade. We do not know of any existing cognitive constraints that would produce these symmetrical results for context effects. It is possible to construct a model that is narrowly constrained to satisfy the QQ equality, but these constraints could also prevent the model from accounting for order effects (see SI Text for two such examples, one based on a model that assumes a probability of repeating the first choice, and another that assumes an anchoring-adjustment process). What is needed is a general theory for question order effects that satisfies the QQ equality constraint. We hope that these findings prompt researchers to look for alternative accounts. In any event, we turn next to the basis for the quantum theory prediction.

Quantum Model of Measurement Order Effects

The discovery of the QQ equality was not an accident. This law was predicted a priori from a quantum probability model of human judgment (5). The model is simple and intuitive, and the derivation for the test is general and parameter free. We begin with a cognitive-process interpretation of the theory and later present it formally. (See ref. 3 for a general introduction to quantum probability applied to cognition and decision.)

The general idea may be stated in the following way. The knowledge that a person has and uses to answer questions can be represented as a very high multidimensional space, H. This space can be described by a set of orthogonal axes (technically termed “basis vectors” below) that is chosen to answer the questions. Many cognitive theories represent knowledge as a vector of feature values, and one can think of the axes in these terms. For example, if features are binary and there are 100 relevant features, then each of the 2^{100} axes can be used to represent a different pattern of ones and zeros where a 1 represents presence and a zero represents absence of that feature (e.g., ref. 8). A person’s beliefs about events are represented by a unit length vector, generally at an oblique angle with respect to these axes. The projection of the belief vector onto an axis can be described as a belief that a feature is present. H does not change with the question asked or with the context in which the question occurs, but the way the knowledge in H is used changes with both factors. For example, not all of the knowledge in H is needed to answer a given question, and the knowledge that is to be considered for answering a given question, A, is a subspace, SA, of H. The knowledge used to answer another question, B, is represented by another subspace, SB, which generally is of different dimensionality and is not necessarily aligned with the axes chosen to describe SA. For example, if H is represented by a cube and SA is the square plane on the bottom of the cube, then SB could be another plane containing the cube’s major diagonal. Finally, the probability of affirming an answer is determined by the square of

![Correlation between Context Effect Pairs](image1)

**Fig. 1.** Empirical demonstration of the QQ equality. Left shows the context effect from one diagonal cell plotted against the context effect from the other cell within the same diagonal. The QQ equality is satisfied when the data points fall on the predicted line with the intercept of 0 and the slope of −1 (shown here). Right shows the ratio (q value size/order effect size) plotted as a function of the size of the order effect, and the QQ equality predicts that, when order effects are large enough to emerge from the noise, then this ratio should start below 1.0 and go to zero as the order effect increases.
the projection of the current belief vector onto the subspace used to answer the question. [One might question the necessity of computing probabilities based on squared length of projections; however, Gleason’s theorem (9) proves that this is the only way to assign an additive probability measure to all subspaces of dimension greater than 2.] It is useful to make this abstract description concrete by considering the following toy example. Imagine a survey respondent who is asked questions about Bill Clinton during the period when Clinton was the US president, such as “Is Clinton a respectable leader?” and “Is Clinton doing a good job as the president?” Suppose for simplicity that only two binary features are used by the respondent to answer these questions: The economy is doing well (yes = 1, no = 0), and a leader should exhibit marital fidelity (yes = 1, no = 0). The four combinations, (11, 10, 01, 00), form a four-dimensional space $H$, spanned by four basis vectors, one for each combination. It is difficult to visualize a four-dimensional space, and so we will assume that nonzero beliefs (that is, 11, 10, and 01) are assigned by the respondent only to the first three of the four basis vectors and zero (that is, 00) is assigned to the fourth basis vector. In Fig. 2, the first three basis vectors are labeled $X$ (corresponding to 11), $Y$ (corresponding to 10), and $Z$ (corresponding to 01). Assume that the answer “yes” to the question “Is Clinton doing a good job as the president?” is satisfied by the property “the economy is doing well,” and so it is represented by the plane $C$ defined by the axes $X$ and $Y$. The orthogonal axis $Z$ represents the answer “no” to this Clinton question. How these subspaces are used depends on the present context, including the way the person is thinking at the time, which determines the weighting on pieces of information in the subspace. The current context and current way of thinking is represented as a “belief vector” of unit length, denoted here as $S$, located at some orientation with respect to the $X$, $Y$, $Z$ axes. One can think of this current belief as defined by the current contents in the person’s short-term memory. In Fig. 2, the initial belief state $S$ is the largest on the 01 combination $Z$. As illustrated in Fig. 2, Upper, the probability of answering “yes” to the question “Is Clinton doing a good job?” is obtained by projecting the current belief vector $S$ down onto the subspace $C$, and squaring its length (which equals 0.33 in this example).

Imagine a similar question is asked about Al Gore: “Would Gore be a good next president instead?” Suppose the answer “yes” to this Gore question is represented by the one-dimensional subspace spanned by the vector $G$ in Fig. 2, which lies at an oblique angle with respect to the subspace $C$. Here, as illustrated by the Clinton and the Gore questions, subspaces used to answer different questions can have different dimensionality and at different orientations with respect to each other. Note that the subspace for the answer “yes” to the Gore question is a ray that is not contained in $C$ (economy is doing well), and it is not aligned with $Z$ (economy is not doing well) either. Psychologically, this represents the idea that the person prefers the answer “yes” to the Gore question when there is uncertainty about the state of the economy. As illustrated in Fig. 2, Lower, the probability of answering “yes” to the Gore question is obtained by projecting the current belief vector $S$ onto $G$ and squaring its length (which equals 0.97 in this example).

Now we come to the heart of the model: context and order effects. As time passes and new information arrives, the content of short-term memory changes, and the belief vector changes accordingly. When two questions immediately follow each other, then after answering the first question, the belief vector that was used to answer the first question changes to match the answer just given. In other words, the belief vector realigns with the current contents of short-term memory (which includes answers to previous questions) and the perspectives that flow from these contents. In geometric terms, the new belief vector used for the second question is simply the projection of the initial belief vector onto the subspace used to answer the first question, normalized to have unit length. This new belief vector is then projected onto the subspace used for answering the second question, and squared to produce the probability of a “yes” response to the second question. Using this process, the probability of the sequence of “yes” answers equals the squared length of the projection produced by first projecting the belief state onto the subspace for answering “yes” to the first question, and then projecting the updated belief vector onto the subspace for answering “yes” to the second question. When the two subspaces lie at oblique angles with respect to each other, the order of answering the questions will change the projections and ultimately the probabilities of the responses, and this is described as “noncommutativity” in quantum theory. This is where a context effect arises.

This process is illustrated in Fig. 2 for the case in which the answers to both questions are “yes.” Fig. 2, Upper shows the process when the Clinton question is asked before the Gore question. The initial state $S$ is first projected onto the plane $C$ and then projected onto the ray $G$, and the squared length of the final projection equals 0.17, which gives the probability of “yes” to Clinton and then “yes” to Gore. Fig. 2, Lower shows the process when the Gore question is asked before the Clinton...
order effects are described as described as the QQ equality in the first part of this article. This particular relation among the observed context/order effects, the one processes. However, the quantum formulation predicts a partic-
might be explained by any number of cognitive theories and no order effects, and the QQ equality holds for trivial reasons.

Up to this point, we have merely presented a “geometric” model description of the way that the quantum formulation will produce context/order effects. The existence of such effects might be explained by any number of cognitive theories and processes. However, the quantum formulation predicts a particular relation among the observed context/order effects, the one described as the QQ equality in the first part of this article. This prediction follows from the “law of reciprocity” often discussed in quantum theory (e.g., ref. 10, p. 34). The essential idea is that the transition from one state to another depends only on the correlation between the states as measured by their inner product. This prediction will be laid out in the formal description to follow.

Now consider the general problem of computing answers to a sequence of questions. Recall that each answer to a question is represented by a subspace (e.g., a ray, a plane, a hyperplane) in an N-dimensional vector space, and so a pair of answers to questions is represented by two different subspaces. Each subspace corresponds to a “projector” that projects state vectors onto the subspace. The probability of agreeing to question A (denoted by \( \mathbf{A} \) and represented by subspace \( \mathbf{A}_0 \)) and then agreeing to question B (denoted by \( \mathbf{B} \) and represented by subspace \( \mathbf{B}_0 \)) is equal to the squared length of the result obtained by sequentially projecting the prior belief state on the subspace of interest to A, and then onto the subspace of interest to B, that is,

\[
p(A \cap B) = |p(A)p(B)_0|^2.
\]

The probability of agreeing to question B and then the belief vector to the subspace \( \mathbf{B}_0 \) is equal to the squared length of the result obtained by sequentially projecting the prior belief state on the subspace for agreeing to B and then on the subspace for agreeing to A, that is,

\[
p(B \cap A) = |p(B)p(A)_0|^2.
\]

If the projectors are commutative (i.e., \( p(A)p(B) = p(B)p(A) \)), then the subspaces are compatible, and no order effects are predicted to occur. If the projectors are noncommutative (i.e., \( p(A)p(B) \neq p(B)p(A) \)), then the subspaces for the two questions are incompatible, and order effects are predicted to occur.

**QQ Equality Derived from the Quantum Model**

If two questions, adjacent to each other, are asked in different orders, then the quantum model of measurement order as described above makes an a priori and parameter-free prediction, namely the QQ equality (see SI Text or ref. 5 or ref. 3 for proofs):

\[
q = |p(A)p(B) + p(A)p(B) - p(B)p(A) - p(B)p(A)|^2.
\]

The first line implies that the two main diagonal cells of the context effect table sum to zero, and the second line implies that the two off-diagonal cells in the context effect table sum to zero (see Table 1 for examples). Intuitively, this means, the probability of having the same response to the two questions should remain invariant across the two question orders; also the probability of having different responses to the two questions should remain invariant across the two question orders. This equality must hold even when context effects produced by the question order occur so that, for example, \( p(A)p(B) \neq p(B)p(A) \) and \( p(A)p(B) \neq p(B)p(A) \). As shown in the proof (SI Text), this equality must hold for any initial belief state and any pair of projectors in any high-dimensional vector space. The QQ equality is still predicted even if there are individual differences in the initial belief state \( \mathbf{S} \), so that it continues to hold when we average across individuals with different belief states (i.e., a mixed state; SI Text).

As we have shown, this precise prediction can be easily tested empirically: If it holds, the difference in observed proportions on the left hand of the QQ equality, defined as \( q \), should not statistically differ from zero as tested by a \( \chi^2 \) test for difference in proportions. As introduced earlier, this \( q \) test was indeed satisfied for the large dataset of 66 national representative surveys on various topics.

Why does the quantum model predict the QQ equality? The proof (SI Text) is based on a fundamental principle of quantum theory called the “law of reciprocity” (10). The probability of transiting from a projection on subspace \( \mathbf{A}_0 \) to a projection on subspace \( \mathbf{B}_0 \) equals the probability of transiting from a projection on subspace \( \mathbf{B}_0 \) to a projection on subspace \( \mathbf{A}_0 \). More formally, for any given state vector \( \mathbf{S} \) and two projectors \( \mathbf{P}_A \) and \( \mathbf{P}_B \), if \( T = \mathbf{P}_A \mathbf{S} \) and \( V = \mathbf{P}_B \mathbf{S} \), then \( |\mathbf{T}|^2 = |\mathbf{V}|^2 \).

A critical assumption underlying the derivation of the QQ equality is that starting from a common state (technically, the state is represented by a density matrix for a mixture of people with individual differences), questions are asked back to back without any information inserted in between so that two successive projections are applied (either \( \mathbf{P}_A \mathbf{P}_B \) or \( \mathbf{P}_B \mathbf{P}_A \)) to the common state. New information presented before or inserted in between questions will change the state in different ways. For example, if new information is presented in between questions, then a change is produced for a transformation \( \mathbf{U}^* \) for one type of information, and a different change is produced by another transformation \( \mathbf{U}^* \) for a different type of information. Rather than comparing \( \mathbf{P}_A \mathbf{P}_B \mathbf{S} \) with \( \mathbf{P}_B \mathbf{P}_A \mathbf{S} \) (as required for the derivation of the QQ equality), we are now comparing \( \mathbf{P}_A \mathbf{U}^* \mathbf{P}_B \mathbf{S} \) with \( \mathbf{P}_B \mathbf{U}^* \mathbf{P}_A \mathbf{S} \), and the equality is no longer expected to hold. (For example, the Rose–Jackson poll in Table 1 violated this condition for our \( q \) test. A more complex quantum model that includes the transformations \( \mathbf{U}^* \) and \( \mathbf{U}^* \) is required for this case.)

The QQ equality prediction from our model depends on the assumption that the belief vector used to answer the second question is the normalized projection used to assign probability to the first question. We have looked at deviations from this assumption: If the new belief vector is assumed to be some proportion of the angular distance between the initial belief vector and the resultant projection, the \( q \) value is not exactly zero; but in all of the cases we have examined, it is very close to zero (SI Text). This suggests that the prediction derived from our quantum model may be even more general than claimed here and may apply to a wider class of cognitive context effects. We leave this possibility open for further research.

**Discussion**

The surprisingly strong evidence for the QQ equality supports the quantum model of measurement order effects. It is part of an accumulating body of research showing that quantum theory can explain a wide range of behavioral findings that are paradoxical.
from a classical probability perspective (11, 12). Very likely, the QQ equality is the strongest form of support because its prediction is not dependent on parameter choices—other applications of quantum theory to human cognition depend on choosing parameter values to best fit data (as do most cognitive models).

This support of the quantum probability approach leaves a question: Can a classical brain give rise to behavior that follows quantum principles (13)? Mathematical physicists have recently provided a plausible account, showing that quantum behavior can emerge when coarse measurements of a classical dynamic system generate incompatible observables that result in unresolvable uncertainty relations and entangled correlations (14, 15). Scientists are still far from understanding how mental states (such as judgments and decisions) emerge from the neural substrates. It is too early to conclude whether or not quantum physics plays a significant role in this emergence (16–19). Regardless, even if the brain’s neural processes operate by classical rules, quantum probability may provide a better description than classical probability for the way human reasons under uncertainty. Applications of both the QQ equality and a wide class of psychological and decision-making tasks (20–24) support this hypothesis, and they share the following conceptual bases: (i) Human judgments, such as attitude judgments, are often not simply read out from memory, but rather, they are constructed from the cognitive state for the question at hand; and (ii) drawing a conclusion from one question changes the context and disturbs the cognitive system, which then (iii) affects the answer to the next question, producing order effects, so that (iv) human judgments do not always obey the commutative rule of Boolean logic. If we replace “human judgments” with “physical measurements” and replace “cognitive system” with “physical system,” then these are exactly the same reasons that led physicists to develop quantum theory in the first place. The QQ equality presented in this paper shows that quantum probability theory, used to explain noncommutativity of measurements in physics, provides a strongly supported prediction for measurement order effects in social and behavioral science.

ACKNOWLEDGMENTS. We thank David Moore for providing the Gallup poll data. We also thank the Pew Research Center for assistance with data acquisition. This work was supported by National Science Foundation Grants 1153846 and 1153726, and Air Force Office of Scientific Research Grant FA9550-12-1-0397.

**Supporting Information**

Wang et al. 10.1073/pnas.1407756111

**SI Text**

**Bounds on q Value.** Consider the Clinton–Gore poll as an example with a pair of questions C and G. Define \( p(C, G_n) \) as the probability of the sequence for saying yes to question C and then yes to question G when C was asked first and G was asked second; define \( p(C, G_n) \) as the probability of the sequence for saying yes to question G and then no to question C when G was asked first and C was asked second; similar definitions are used for the other sequences. The context effect table contains four context effects, which can be defined as shown in Table S1.

The context effects must sum to zero because (\( x + y + z + w = p(C, G_n) = p(C, G_n) + p(C, G_n) - p(G, C_n) + p(G, C_n) - p(C, C_n) = 1 - 1 = 0 \) so that (\( x + w ) + (y + z) = 0 \) → (\( x + w ) = q \).}

The set of all context effects forms a 3D pyramid defined by all \( x, y, z, w \) such that \( x + w + y + z = 0, 1 - p(C, G_n) \geq x \geq 0 - p(C, G_n) \), \( 1 - p(C, G_n) \geq y \geq 0 - p(C, G_n) \), \( 1 - p(C, G_n) \geq z \geq 0 - p(C, G_n) \), and \( 1 - p(C, G_n) \geq w \geq 0 - p(C, G_n) \). The set of context effects that satisfy the quantum question (QQ) equality forms a triangular plane that lies inside this pyramid defined by adding one more constraint \( q = (y + z) = - (x + w) = 0 \). Therefore, the set of context effects that satisfy the QQ equality is a very tiny portion of the set of all possible context effects.

By our definitions of context effects \( x, y, z, w \) above, we have \( p(C, G_n) = p(C, G_n) - y \geq 0 \rightarrow p(C, G_n) - y \geq 0 \), and \( p(C, G_n) - y \geq 0 \rightarrow p(C, G_n) - y \geq 0 \). Similarly, we also have \( p(C, G_n) = p(C, G_n) - x \geq 0 \rightarrow p(C, G_n) - x \geq 0 \), and \( p(C, G_n) = p(C, G_n) - w \geq 0 \rightarrow p(C, G_n) - w \geq 0 \); so, \( p(C, G_n) + p(C, G_n) \geq - (x + w) + w = q \). Therefore, we obtain the bound \( - [p(C, G_n) + p(C, G_n)] \leq q \leq [p(C, G_n) + p(C, G_n)] \). This bound can be much larger than the possible \( q \) value, and we can obtain a tighter bound as described next. We do not use this large bound in the manuscript, and instead use the tighter bound described next.

Order effects are often defined by the difference marginal proportions produced by each order. In particular, define \( Cy = p(C, C_n) + p(C, C_n) \) as the order effect for yes to Clinton, and define \( Gx = p(G, G_n) + p(G, G_n) \) as the order effect for yes to Gore. Likewise, \( Cn = p(C, C_n) + p(C, C_n) \) as the order effect for no to Clinton and \( Gn = p(G, G_n) + p(G, G_n) \) as the order effect for no to Gore. We could use the sum \( Cy + Gx = y - z \) to define the total order effect. In general, however, the answers are arbitrary labels, and thus we could just as well take the sum \( Cy + Gn = x - w \). This provides two different ways to define the order effect. Summing in each of these ways is problematic, because each of these sums could be zero, which would misrepresent the fact that there is indeed an order effect. Hence, to measure the order effect size, we eliminate the sign. This produces two order effect sizes, \( |z| + |y| \) and \( |w| + |x| \), but one could be small or even zero, and the other could be much larger. Therefore, we define the size of the order effect as the maximum of \( |z| + |y| \) versus \( |w| + |x| \). This also provides a meaningful comparison with the size of our \( q \) values. On the one hand, if \( |z| + |y| \) is maximum, we can compare this to the magnitude \( |q| = |y + z| \), which must be less than or equal; if the order effect \( |w| + |x| \) is maximum, we can compare this to \( -|q| = |w + x| \), which again can be less than or equal.

**QQ Equality, Order Effects, and the Nature of the Questions.** According to our quantum model of measurement order effects, the relation between two questions determines the order effect, but not the QQ equality. On the one hand, QQ equality holds regardless the two questions’ relation and the initial belief/attitude state. It is in this sense, the \( q \) test is an a priori and parameter-free test. See latter section in this SI Text for the proof of QQ equality. On the other hand, depending on the question relations and the initial belief/attitude state, the model predicts order effects to vary in their directions and sizes. See ref. 1, table 2, for a summary of the predictions on the order effects.

When two questions are completely unrelated, either people generally do not expect order effects, or if the order effects do occur, they do not find the reasons interesting. Thus, the empirical datasets included in the current study are question pairs that are more or less related (e.g., evaluation on two politicians; religious beliefs and political attitude).

Also see the following discussion from ref. 1 (p. 705) for more detailed discussion on the nature of the questions when we expect question order effects:

According to the QQ model, two questions may be treated as either compatible or incompatible measurements, and order effects only occur for the latter type. Two questions are incompatible if the projectors for the questions are not defined by a common basis. Inability to form a common basis for different questions is expected to occur when the questions are new or unusual, so that an answer must be constructed online. However, if a person has a great deal of experience with a combination, then the person may have sufficient knowledge to form a compatible representation that can answer both questions using a common basis. See Schulkin (2008) for further discussions of this type of cognitive adaption with experience. Therefore, order effects are expected to occur for pairs of incompatible questions, especially uncommon pairs, which must be (partially) constructed on the spot to answer them. Our constructionist view of belief, attitude, and intention is consistent with many others (e.g., Feldman & Lynch, 1988; Schwarz, 2007). From this view, because of cognitive economy, belief, attitude, and intention are not stored in memory as properties, but instead, are constructed while being needed. The retrieved information for one question affects the context for the construction process and influences the subsequent response. Many researchers have investigated the effects of measurements themselves on measured cognition, such as mere measurement effects, self-generated validity theory, and reasons theory (Chandon et al., 2005; Dholakia & Morwitz, 2002; Fitzsimons & Morwitz, 1996; Feldman & Lynch, 1988; Janiszewski & Chandon, 2007). The proposed QQ model attempts to specify and formalize these theories.

**Description of the Datasets.** The 72 datasets included in the current analysis are 70 field experiments (national surveys) and two laboratory experiments. All of the national surveys recruited national representative samples in the United States, and the laboratory experiments recruited college undergraduate students. In all of the experiments, a randomly selected half of the participants were asked the pair of questions, A and B, in the order of A-then-B. The other half answered exactly the same questions but in the opposite order: B-then-A. The two questions were adjacent to each other. The questions provided binary choices. The datasets are described below.

i) Moore (2) reviewed four types of questions order effects and provided an example dataset from Gallup polls for each type. We obtained the data from him. One of the datasets reviewed by him is excluded from the current analysis because the two questions were not asked back to back, and there was additional context information for one of the questions. According to our quantum model, the QQ equality was predicted not to hold and indeed it did not. The
remaining three datasets from Moore (2) are included in the current analysis. See Table S2.

ii) The current analysis also included a dataset from ref. 3. Almost none of the published studies on question order effects reported sequential probabilities of the questions (cross-tabulation of the portions of responses to the two questions by question order), but the article by Schuman et al. (3) included them, so we are able to use their data to test the QQ equality. The data were collected through Survey Research Center (SRC) national monthly telephone surveys. See Table S3.

iii) Two datasets were collected from between-subjects laboratory experiments by the authors. Undergraduate volunteers were recruited from a large Midwestern university in the United States. One experiment was designed to replicate the Gallup white/black racial hostility study reported by Moore (2). The findings indeed replicated the results found by Gallup in 1996, with a larger effect size. The other experiment was designed to replicate a survey field experiment conducted by Wilson et al. (4). See Table S4.

iv) Finally, we obtained from the Pew Research Center all its surveys that included question order manipulation between two questions during 2001–2011 (10 y) whenever the data are available. See Table S5.

The entire dataset used in the current paper, which contains all of the data from the 72 studies as described, is available in a single file called QuestOrdData.txt. The data file is available from the authors by an e-mail request. The file contains a matrix of 72 rows and 8 columns. The eight columns give the frequencies for AyBy, AyBn, AnBn, ByAy, ByAn, BnAy, BnAn, respectively. Each row is a study. The rows are the following studies, respectively: the 66 Pew studies on various topics, the 3 Gallup studies (Clinton–Gore, Dole–Gingrich, black–white), the two laboratory experiments (black–white, affirmative action), and finally, the abortion questions, as described above.

χ² Tests Used in Fig. 1 of Article. First, we present the χ² test for order effects. Define \( n_{NY} \) as the frequency of saying “of category C when G was asked first and saying “no” to question G when G was asked second, and the other combinations of answers are defined similarly. Define \( n=n_{YY}+n_{YN}+n_{NY}+n_{NN} \). Define \( m_{YN} \) as the frequency of “yes” to question G when G was asked first and “no” to question C when C was asked second, and the other combinations of answers are defined similarly. Define \( n=m_{YY}+m_{YN}+m_{NY}+m_{NN} \). The log-likelihood for the unconstrained model that allows order effects is defined by the following:

\[
G_U = [n_{YY} \cdot \ln(n_{YY}/n) + n_{YN} \cdot \ln(n_{YN}/n) + n_{NY} \cdot \ln(n_{NY}/n) + n_{NN} \cdot \ln(n_{NN}/n) + n_{MY} \cdot \ln(n_{MY}/m) + m_{YN} \cdot \ln(m_{YN}/m) + m_{NY} \cdot \ln(m_{NY}/m) + m_{NN} \cdot \ln(m_{NN}/m)].
\]  

[S1a]

The log-likelihood for the constrained model that assumes no order effects is defined by the following:

\[
G_C = [(n_{YY} + m_{YY}) \cdot \ln((n_{YY} + m_{YY})/(n + m)) + (n_{YN} + m_{YN}) \cdot \ln((n_{YN} + m_{YN})/(n + m)) + (n_{NY} + m_{NY}) \cdot \ln((n_{NY} + m_{NY})/(n + m)) + (n_{NN} + m_{NN}) \cdot \ln((n_{NN} + m_{NN})/(n + m))].
\]  

[S1b]

The \( \chi^2 \) statistic is defined by the difference \( \chi^2 = -2 \cdot (G_C - G_U) \). The unconstrained model involves \((4 - 1) + (4 - 1) = 6\) free parameters and the constrained model involves \(4 - 1 = 3\) free parameters, and so the \( \chi^2 \) statistic has \( df = 3 \).

Next, we define the \( \chi^2 \) test for the QQ equality. The log-likelihood for the unconstrained model is defined as follows:

\[
G_U = [(n_{YY} + n_{YN}) \cdot \ln((n_{YY} + n_{YN})/n) + (n_{YN} + n_{NN}) \cdot \ln((n_{YN} + n_{NN})/n) + (n_{NY} + n_{NN}) \cdot \ln((n_{NY} + n_{NN})/m) + (n_{YY} + n_{MN}) \cdot \ln((n_{YY} + n_{MN})/m)].
\]

[S2a]

The log-likelihood for the model constrained by the QQ equality equals the following:

\[
G_C = [(n_{YY} + n_{YN} + n_{YN} + m_{YN}) \cdot \ln((n_{YY} + n_{YN} + m_{YN} + m_{YN})/(n + m)) + (n_{YN} + n_{NN} + n_{NN} + m_{NN}) \cdot \ln((n_{YN} + n_{NN} + m_{YN} + m_{MN})/(n + m))].
\]

[S2b]

The \( \chi^2 \) statistic is defined by the difference \( \chi^2 = -2 \cdot (G_C - G_U) \). The unconstrained model involves \((2 - 1) + (2 - 1) = 2\) free parameters and the constrained model involves \(2 - 1 = 1\) free parameter, and so the \( \chi^2 \) statistic has \( df = 1 \).

χ² Tests Used to Test Order Effects and q Values in the Article. First, we describe the \( \chi^2 \) test for order effects. The \( \chi^2 \) statistic for testing an order effect for each of the 66 Pew datasets (not specifically picked to have order effects) was computed using Eqs. S1a and S1b defined above, producing 66 observed \( \chi^2 \) values. If the null hypothesis is correct, these should be distributed according to a central \( \chi^2 \) distribution with \( df = 3 \). Ten categories were constructed by computing the nine equally spaced category bounds. These category bounds divide the expected frequency distribution (under the null hypotheses) into 10 equally likely categories, with an equal expected frequency within each of the 10 categories. Then frequency of the 66 observed \( \chi^2 \) values were counted for each category. Denote \( f_i \) as the observed frequency for category \( i = 1,10 \). The log-likelihood for the unconstrained model equals the following:

\[
G_U = \sum_{i} f_i \cdot \ln(f_i/66).
\]

[S3a]

The log-likelihood for the expected frequencies according to the null hypothesis equals the following:

\[
G_C = \sum_{i} f_i \cdot \ln(6.6/66).
\]

[S3b]

The \( \chi^2 \) statistic is defined by the difference \( \chi^2 = -2 \cdot (G_C - G_U) \). The unconstrained model involves \(10 - 1 = 9\) free parameters and the constrained model has no free parameters, and so the \( \chi^2 \) statistic has \( df = 9 \).

Next, we describe the \( \chi^2 \) test for the QQ equality. The \( \chi^2 \) statistic for testing the QQ equality for each of the 66 Pew datasets was computed using Eqs. S2a and S2b defined above, producing 66 observed \( \chi^2 \) values. If the null hypothesis is correct, these should be distributed according to a \( \chi^2 \) distribution with \( df = 1 \). Ten categories were constructed by computing the nine equally spaced category bounds. These category bounds divide the expected frequency distribution (under the null hypotheses) into 10 equally likely categories. Then, frequency of the 66 observed \( \chi^2 \) values was counted for each category. Denote \( f_i \) as the observed frequency for category \( i = 1,10 \). Then Eqs. S3a and S3b were used to compute the log-likelihoods of the unconstrained and constrained models. Once again, the \( \chi^2 \) statistic is defined by the
difference $\chi^2 = -2$. (GC - Gu). The unconstrained model involves 10 − 1 = 9 free parameters and the constrained model has no free parameters, and so the $\chi^2$ statistic has df = 9.

Fig. S1 shows two quantile–quantile plots: Fig. S1, Upper is for the order effect and the bottom is for the q value. The horizontal axis represents the quantile predicted by the $\chi^2$ distribution under the null hypothesis, and the vertical axis represents the observed quantile for the $\chi^2$ statistic. If the null hypothesis is correct, then the points should all fall on the unit slope line. As can be seen in Fig. S1, Upper for order effects, the points fall well above the prediction line, and a $\chi^2$ distribution test produced a significant deviation from the null hypothesis ($p = 0.0004$). As shown by Fig. S1, Lower for the q values, all of the points fall right along the prediction line, and a $\chi^2$ distribution test indicates no significant deviation from the null hypothesis ($p = 0.4625$).

Alternative Models. Wang and Busemeyer (1) have shown that the $\chi^2$ statistic has df

$$p + \frac{1}{2}$$

Wang and Busemeyer (1) have shown that

$$p(\text{AyBy}) = p(\text{ByAy})$$

and the difference between marginal probabilities $p(\text{Ay}) - p(\text{By})$. As can be seen, empirically there is little or no correlation ($r = 0.04$), although the predicted correlation equals 1. In particular, the repeat-choice model predicts that $p(\text{AyBy}) - p(\text{ByAy})$ must be smaller than $|p(\text{Ay}) - p(\text{By})|$. However, we find that 56 of the 72 datasets produce a context effect that is bigger than the difference between marginal probabilities.

There is an additional problem faced by the repeat-choice model. The model can be constrained to satisfy the QQ equality, but it fails to account for the order effects. It predicts that context effects are completely determined by the differences in marginal choice probabilities. We computed the $\chi^2$ statistics for all of the 72 datasets. The distribution of the 72 $\chi^2$ statistics obtained by comparing the constrained repeat-choice model (5 parameters) with the unconstrained model (6 parameters) is significantly different from that expected by a central $\chi^2$ distribution with 1 df (the 72 $\chi^2$ statistics were divided into 10 equally spaced categories). The $\chi^2$ statistics equals 168 (df = 9), $p < 0.0001$.

Anchoring-adjustment model. Define the observed probabilities for the A-then-B order as $p_{\text{AB}}(\text{AyBy})$, $p_{\text{AB}}(\text{AyBn})$, $p_{\text{AB}}(\text{AnBy})$, $p_{\text{AB}}(\text{AnBn})$. These are constructed from the implicit joint probability plus a bias as follows: $p_{\text{AB}}(\text{AyBy}) = p(\text{Ay})/p(\text{AyBy}) + b_{\text{AB}}$, $p_{\text{AB}}(\text{AyBn}) = p(\text{Ay})/p(\text{AyBn}) + b_{\text{AB}}$, $p_{\text{AB}}(\text{AnBy}) = p(\text{An})/p(\text{AnBy}) + b_{\text{AB}}$, $p_{\text{AB}}(\text{AnBn}) = p(\text{An})/p(\text{AnBn}) - b_{\text{AB}}$. The bias $b_{\text{AB}}$ can be positive or negative but it must be constrained so that all of the observed probabilities are in fact probabilities (i.e., real numbers between zero and 1).

We can estimate each bias directly from each two by two table of responses (see Fig. 1 in the main article for examples) using the mathematical derivation for each bias from the model: $b_{\text{AB}} = [p(\text{By}) - p(\text{ByAy})]/[p(\text{AyBy}) - p(\text{AyBn})]$, $b_{\text{BA}} = [p(\text{Ay}) - p(\text{AyBn})]/[p(\text{AyBy}) - p(\text{AyBn})]$.

Similarly, define the observed probabilities for the B-then-A order as $p_{\text{BA}}(\text{ByAy})$, $p_{\text{BA}}(\text{ByBn})$, $p_{\text{BA}}(\text{BnAy})$, $p_{\text{BA}}(\text{BnBn})$. These are constructed from the implicit joint probability plus a bias as follows: $p_{\text{BA}}(\text{ByAy}) = p(\text{By})/p(\text{ByAy}) + b_{\text{BA}}$, $p_{\text{BA}}(\text{ByBn}) = p(\text{By})/p(\text{ByBn}) + b_{\text{BA}}$, $p_{\text{BA}}(\text{BnAy}) = p(\text{Bn})/p(\text{BnAy}) - b_{\text{BA}}$, $p_{\text{BA}}(\text{BnBn}) = p(\text{Bn})/p(\text{BnBn}) - b_{\text{BA}}$. Again, the bias can be positive or negative, but the bias $b_{\text{BA}}$ must be constrained so that the observed probabilities are in fact probabilities.

If the bias for the B-then-A order differs from the bias for the B-then-A order, then this model does not satisfy the QQ equality. The model with unequal bias has five free parameters (leaving only 1 df to test the theory). However, if we force the two biases to be equal, $b_{\text{BA}} = b_{\text{AB}}$, then this model does satisfy the QQ equality. This constrained model has four free parameters (leaving 2 df’s to test the model).

We can estimate each bias directly from each two by two table of responses (see Fig. 1 in the main article for examples) using the mathematical derivation for each bias from the model: $b_{\text{AB}} = [p(\text{By}) - p(\text{ByAy})]/[p(\text{AyBy}) - p(\text{AyBn})]$, $b_{\text{BA}} = [p(\text{Ay}) - p(\text{AyBn})]/[p(\text{AyBy}) - p(\text{AyBn})]$.

Fig. S2 below shows one bias plotted as a function of the other computed from the 72 datasets used in the paper. If the biases are equal, then the data points should all fall along a line with zero intercept and unit slope. As can be seen from Fig. S3 below, the observed correlation between the two biases is almost zero, which does not support the equal bias assumption.

An additional problem faced by the anchoring-adjustment type of models is the following: If the two biases for the two orders are constrained to be equal so that the model satisfies the QQ equality, then the model fails to account for the order effects. We computed the $\chi^2$ statistics for all of the 73 datasets. The distribution of $\chi^2$ statistics from the constrained anchoring-adjustment model is significantly different from the distribution expected under the null hypothesis (the null state that there is no difference between the constrained anchoring-adjustment model and the saturated model).

The $\chi^2$ statistics equals 39 (df = 9), $p < 0.0001$.

Proof of the QQ Equality. Here, we briefly introduce the basic axioms of quantum theory and then derive the QQ equality. We use the Dirac bracket notation so that $\langle S\mid T \rangle$ represents the inner
product between two vectors. According to quantum theory, events are represented as subspaces of a Hilbert space. Corresponding to each event A there is a orthogonal projector $P_A$. The state of a quantum system is represented by a unit length vector $S$ within the Hilbert space [Busemeyer and Bruza (ref. 5, appendix C) use a more general density matrix to represent the state]. The probability of event $A$ equals the squared length of the projection $p(A) = |P_A S|^2$. If event $A$ is observed, then the state is updated according to Lüder’s rule $S' = P_A S / ||P_A S||$.

Define $S$ as the initial state (see ref. 5, appendix C, for proof using density matrix representation of state). Denote the projector for saying yes to question $C$ as $P_C$ and denote $P_G$ as the projector for saying yes to question $G$. We start by expanding the probability for answering “yes” to question $C$:

$$
|P_C \cdot S|^2 = |P_C \cdot S|^2 = |P_C \cdot (P_G \cdot (I - P_G)) \cdot S|^2 \\
= |P_C \cdot P_G \cdot S + P_C \cdot (I - P_G) \cdot S|^2 \\
= |P_C \cdot P_G \cdot S|^2 + |P_C \cdot (I - P_G) \cdot S|^2 + \langle S | (I - P_G) \cdot P_C \cdot (I - P_G) | S \rangle \\
= |P_C \cdot P_G \cdot S|^2 + |P_C \cdot (I - P_G) \cdot S|^2 + \langle S | P_G \cdot P_C \cdot (I - P_G) | S \rangle \\
= |P_C \cdot P_G \cdot S|^2 + |P_C \cdot (I - P_G) \cdot S|^2 + 2 \cdot \text{Re} \left[ \langle S | P_G \cdot P_C \cdot (I - P_G) | S \rangle \right]
$$

and the latter follows from the idempotent property of projectors. (The symbol $x^*$ used in the above derivation refers to the complex conjugate of $x$.) Define the total probability to say yes to question $C$ when $G$ was asked first as follows:

$$
T_P = |P_C \cdot S|^2 = |P_C \cdot P_G \cdot S|^2 + |P_C \cdot (I - P_G) \cdot S|^2 + \langle S | P_G \cdot P_C \cdot (I - P_G) | S \rangle
$$

An order effect for question $C$ when $G$ was asked first is expressed as follows:

$$
I_C = T_P - |P_C \cdot S|^2 = -2 \cdot \text{Re} \left[ \langle S | P_G \cdot P_C \cdot (I - P_G) | S \rangle \right].
$$

Immediately, we see that if $P_G$ and $P_C$ commute so that $P_G \cdot P_C = P_C \cdot P_G$ then $P_G \cdot P_C \cdot (I - P_G) = P_C \cdot P_G \cdot (I - P_G) = 0$ and we predict no order effect. Thus, noncommuting projectors are a necessary condition for order effects. Now let us reexamine the following:

$$
I_C = T_P - |P_C \cdot S|^2 = -2 \cdot \text{Re} \left[ \langle S | P_G \cdot P_C \cdot (I - P_G) | S \rangle \right].
$$

Similarly, the order effect for question $C$ when $G$ was asked first equals the following:

$$
I_G = T_P - |P_G \cdot S|^2 = 2 \cdot p(G \cdot C) - 2 \cdot \text{Re} \left[ \langle S | P_G \cdot P_C \cdot (I - P_G) | S \rangle \right].
$$

These two order effects share the same term, $2 \cdot \text{Re} \left[ \langle S | P_G \cdot P_C \cdot S | S \rangle \right]$. These two terms must be equal because of the mathematical property of inner products:

$$
\langle \psi | \phi \rangle = | \langle \psi | \phi \rangle | = | \langle \phi | \psi \rangle |.
$$

The law of reciprocity requires that the transition probability from state $y$ to $x$ be equal to the transition probability from $x$ to $y$ because the transition probabilities are determined by the inner products: $p(x|y) = | \langle x | y \rangle |^2 = p(y|x)$.

The two order effects, $I_C$ and $I_G$, together imply the following relation:

$$
0 = \langle 2 \cdot p(G \cdot C) - 2 \cdot p(C \cdot G) \rangle - I_G = \langle 2 \cdot p(G \cdot C) - p(C \cdot G) - p(C \cdot G) + p(C \cdot G) \rangle - \langle 2 \cdot p(C \cdot G) - p(C \cdot G) + p(C \cdot G) \rangle
$$

As noted above, the critical property for this prediction is the quantum law of reciprocity, which requires the transition amplitude $(T | S)$ from $S \rightarrow T$ to be the conjugate of the transition amplitude $(S | T)$ from $T \rightarrow S$ so that $(S | T)^* = (T | S)$. Busemeyer and Bruza (ref. 5, appendix C) present a more general proof using a density matrix to represent the state. The density matrix allows for the state to represent a population with individual differences in the pure state $S$. This equality also appears as one of the axioms in the axiomatic formulation of quantum theory by Niestegge (6).

**Partial Revision Models.** Quantum theory assumes that, after a first choice is made, the revised belief state is formed from the initial belief state by projecting the initial belief state onto the subspace for an answer to a question and normalizing the projection to form the revised belief state. More generally, maybe the revision can be “partial” in the sense that the revision is formed by moving some proportion of the belief from the initial belief state to the final projection. We examined the QQ equality under this more general, relaxed assumption for the revision using computer simulation. We examine proportions for revisions of 0, 0.3, 0.5, 0.6, 0.7, and 1.0. (Of course, proportions at the extremes of 0 or 1 exactly satisfy the QQ equality.) We examined spaces ranging in dimension from 2 to 10. We examined two completely different types of unitary transformations that relate the subspaces for the answers to the two questions. We examined subspaces ranging from 1 to 5 in dimension (limited by the size of the entire space). We examined 10,000 random initial states for each of the combinations above. All of the simulations produced QQ equalities of size less than 0.015 and most less than 0.01. So the QQ equality is not exactly satisfied but approximately satisfied for this relaxed version of the theory. The MATLAB program used to perform these simulations is provided below:
% Partial revision model: revised state moves to position between initial state and normalized projection
clc

clear

N = 10     % number of dimensions for H
QV = [];
a = .6;    % proportion of movement. a=0 -> stay at initial state
b = 1-a;   % a=1 -> move to projection
Trep = 10000;    % number of random selections of initial state

for rep = 1:Trep

    S = rand(N,1);         % random initial state
    S = S./sqrt(S'*S);
    % S = ones(N,1);        % fixed initial state uniform
    % S = S./sqrt(N);
    H1 = [1 -1; 1 1]./sqrt(2);    % Hadamard gate
    % H = kron(H1,H1);
    % U = kron(H1,H);

    H = zeros(N);                       % Hamiltonian
    H(2:N,1:N-1) = diag(ones(N-1,1));
    H = H + H';
    H = H + diag(1*(1:N));
    U = expm(-1i*H);                  % Unitary

    na = 1;                                    % dim of A subspace
    PA = diag([ones(na,1); zeros(N-na,1)]);    % projector for A
    PnA = eye(N) - PA;                         % projector for not A

    SA = PA*S; SA = SA./sqrt(SA'*SA);          % normalized projection A
    Sa = a*SA + b*S;                           % revised state
    Sa = Sa./sqrt(Sa'*Sa);                     % re-normalized
    SnA = PnA*S; SnA = SnA./sqrt(SnA'*SnA);    % normalized projection B
    Sna = a*SnA + b*S;
    Sna = Sna./sqrt(Sna'*Sna);

    % Wang et al. www.pnas.org/cgi/content/short/1407756111

nb = 2; % dim of B subspace
% PB = diag([zeros(N-nb,1); ones(nb,1)]);
PnB = eye(N) - PB;
% Pb = a*PB + b*eye(N);
% Pnb = a*PnB + b*eye(N);
% Sb = Pb*S; Sb = Sb./sqrt(Sb'*Sb);
% Snb = PnB*S; Snb = Snb./sqrt(Snb'*Snb);
SB = PB*S; SB = SB./sqrt(SB'*SB);
Sb = a*Sb + b*S;
SnB = PnB*S; SnB = SnB./sqrt(SnB'*SnB);
Snb = a*SnB + b*S;

pAtB = (S'*PA*S)*(Sa'*PB*Sa); % prob A then B
pAtnB = (S'*PA*S)*(Sa'*PnB*Sa);
pnAtB = (S'*PnA*S)*(Sna'*PB*Sna);
pnAtnB = (S'*PnA*S)*(Sna'*PnB*Sna);
pAtB+pAtnB+pnAtB+pnAtnB;

pBtA = (S'*PB*S)*(Sb'*PA*Sb); % prob B then A
pBtA = (S'*PB*S)*(Sb'*PA*Sb);
pnBtA = (S'*PnB*S)*(Snb'*PA*SnB);
pnBtA = (S'*PnB*S)*(Snb'*PA*SnB);
pBtA+pBtA+pnBtA+pnBtA;

x = pAtB - pBtA; % order effects
w = pnAtnB - pnBtA;
y = pAtnB - pnBtA;
z = pnAtB - pBtA;

Q1 = x+w; % Q value
Q2 = y+z;
QV = [ QV Q1 ];
end
max(abs(QV))
Fig. S1. Quantile-quantile plots of $\chi^2$ statistics. Upper shows order effects, and Lower shows $q$ values. In each panel, the horizontal axis represents the quantile predicted by the $\chi^2$ distribution under the null hypothesis, and the vertical axis represents the observed quantile for the $\chi^2$ statistic. The green line is the prediction for the null hypothesis; the blue circles are the observed values. Deviations indicate departures from the null hypothesis.

Fig. S2. Correlation between the context effect and the difference between marginal probabilities in the 72 datasets is very small. The blue circles represent the pairs (context effect, difference between marginal probabilities) for each of the 72 studies. The scatter plot shows the relationship between the context effect and the difference between marginal probabilities in each of the 72 studies. The repeat-choice model predicts a correlation equal to 1.
Prediction of an anchor–adjustment model

$r = -0.003$

Bias from BA

Bias From AB

Fig. S3. Correlation between the two biases in the anchoring-adjustment model, as estimated from the 72 datasets, is almost zero. The green line is the regression line with zero intercept and unit slope predicted by the anchor-adjustment model. The blue circles represent the pairs of estimated biases from the AB order and from the BA order (derived from the anchor-adjustment model) obtained from each of the 72 studies. The scatter plot illustrates correlation between the bias in the AB and BA orders. The correlation is predicted to be 1 by the anchor-adjustment model.

Table S1. The context effects in two question sequences

<table>
<thead>
<tr>
<th>The first order effect</th>
<th>The second order effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = p(G, C_y) - p(C, G_y)$</td>
<td>$y = p(G, C_y) - p(C, G_y)$</td>
</tr>
<tr>
<td>$z = p(G, C_n) - p(C, G_n)$</td>
<td>$w = p(G, C_n) - p(C, G_n)$</td>
</tr>
</tbody>
</table>

Table S2. Three Gallup field experiments reported by Moore (1)

<table>
<thead>
<tr>
<th>Gallup surveys</th>
<th>N</th>
<th>Question A</th>
<th>Question B</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 27–29, 1995</td>
<td>1,016</td>
<td>Whether you think Newt Gingrich is honest and trustworthy? (Yes/No)</td>
<td>Whether you think Bob Dole is honest and trustworthy? (Yes/No)</td>
</tr>
<tr>
<td>June 27–30, 1996</td>
<td>1,005</td>
<td>Do you think that only a few or many white people dislike black people? (A few/Many)</td>
<td>Do you think that only a few or many black people dislike white people? (A few/Many)</td>
</tr>
<tr>
<td>September 6–7, 1997</td>
<td>1,003</td>
<td>Do you generally think Bill Clinton is honest and trustworthy? (Yes/No)</td>
<td>Do you generally think Al Gore is honest and trustworthy? (Yes/No)</td>
</tr>
</tbody>
</table>


Table S3. An SRC field experiments reported by Schuman et al. (1)

<table>
<thead>
<tr>
<th>Gallup surveys</th>
<th>N</th>
<th>Question A</th>
<th>Question B</th>
</tr>
</thead>
<tbody>
<tr>
<td>March–August 1979</td>
<td>651</td>
<td>Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children? (Yes/No)</td>
<td>Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of serious defect in the baby? (Yes/No)</td>
</tr>
</tbody>
</table>

Table S4. Two laboratory experiments

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>Question A</th>
<th>Question B</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012–2013</td>
<td>116</td>
<td>Do you think that only a few or many white people dislike black people? (A few/Many)</td>
<td>Do you think that only a few or many black people dislike white people? (A few/Many)</td>
</tr>
<tr>
<td>2012–2013</td>
<td>118</td>
<td>Do you generally favor or oppose affirmative action (AA) programs for racial minorities? (Yes/No)</td>
<td>Do you generally favor or oppose affirmative action (AA) programs for women? (Yes/No)</td>
</tr>
</tbody>
</table>

Table S5. Sixty-six Pew field experiments during 2001–2011

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>Question A</th>
<th>Question B</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001–2011, 26 studies</td>
<td>Range: 815–3,006</td>
<td>Do you approve or disapprove of the way Bush/Obama is handling his job as President? (Approve/Disapprove)</td>
<td>All in all, are you satisfied or dissatisfied with the way things are going in this country today? (Satisfied/Dissatisfied)</td>
</tr>
<tr>
<td></td>
<td>M  = 1,644</td>
<td>SD = 422.24</td>
<td></td>
</tr>
<tr>
<td>2001–2011, 15 studies</td>
<td>Range: 1,308–1,802</td>
<td>Do you approve or disapprove of the job the Republican leaders in Congress are doing? (Approve/Disapprove)</td>
<td>Do you approve or disapprove of the job the Democratic leaders in Congress are doing? (Approve/Disapprove)</td>
</tr>
<tr>
<td></td>
<td>M  = 1,504</td>
<td>SD = 95.74</td>
<td></td>
</tr>
<tr>
<td>2001–2011, 25 studies</td>
<td>Range: 1,001–2,509</td>
<td>Most pair of questions was asked once. Several were asked twice. Questions cover a large variety of topics, from religious beliefs to support to economic policy.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M  = 1,612.48</td>
<td>SD = 328.28</td>
<td></td>
</tr>
</tbody>
</table>

M, mean.

For example, a pair of questions asked, “Do you approve or disapprove of the way Barack Obama handled the problems faced by US automakers like General Motors and Chrysler?” and “Do you approve or disapprove of the way Barack Obama handled the problems faced by major US banks and financial institutions?” (Approve/Disapprove)

Another example is the following:

“In the past presidential election campaign, do you think Barack Obama/John McCain has been too personally critical of John McCain/Barack Obama, or not?” (Too personally critical/Not too personally critical)