

Frequency, Probability, and Prediction: Easy Solutions to Cognitive Illusions?

Dale Griffin & Roger Buehler

Cognitive Psychology, **38**, 48-78 (1999)

Introduction

- Power of particular (Kahneman, 1998)
 - 複数の事例の集約よりも、一つ的事例について考える効果
 - power of particular は、以下の 2 つの心理現象を包括する
 - ・ Redelmeier & Tversky (1992)
 - 個別の事例は、複数の事例の集約よりも、共感や情緒反応を引き起こす
 - ・ Tversky & Kahneman (1983)
 - 個別の事例を用いる場合、決定論的な特定の推論プロセスを要する
 - 集約された情報を用いる場合、集約された事例の特徴と、個別の事例の特徴との関係について比較する推論プロセスを要する
- 頻度予測
 - Griffin & Tversky (1992), Tversky & Koehler (1994)
 - ・ 個別の事例について予測を行うとき、人は、個別の事例に関する情報にだけ注目し、その事例が広範囲なカテゴリーに関連する可能性を無視する
 - Cosmides & Tooby (1996), Gigerenzer (1991), Griffin & Tversky (1992)
 - ・ 複数の事例に基づく頻度予測は、広範囲な事例から適切な証拠を想起させる
 - Kahneman & Tversky (1979, 1982a)
 - ・ Inside perspective
 - 個別の特徴に焦点を当てること
 - これにより、特定のプランニングが行われ、他の情報は無視される
 - ・ Outside perspective
 - 複数の事例の要約された特徴に焦点を当てること
 - Kahneman & Tversky (1982b), Tversky & Kahneman (1983)
 - ・ 集約された頻度情報は、人間を統計的に、拡張的に思考させる
 - Gigerenzer (1991)
 - ・ 頻度に関する問題に直面したとき、人は、規範的な判断や推論を行うが、確率に関する問題では、行われない
- 研究の目的
 - 事例の集約に基づく頻度予測が、確率的予測にみられる認知的錯覚 (cognitive illusion) をどのように防ぐのか検討する

- 3つの認知的錯覚
 - ・ planning fallacy - 楽観的予測
 - ・ overconfidence in knowledge - 選択した結果が正解であるという予測
 - ・ base-rate neglect - 直観的予測
- Planning fallacy
 - Planning fallacy (Kahneman & Tversky, 1979, 1982a)
 - ・ 課題遂行時間の予測を行う際、最も都合の良い時間の見積もりを行う傾向
 - 過去に類似した課題を行い、その時の課題遂行時間が見積もり時間を超過していることに気づいていても、最も都合の良い時間の見積もりを行う
 - 複数の課題に対する予測を行う場合、人は、過去の複数の事例に注意するようになる (Outside perspective を持つ) のかもしれない

Experiment 1

- 複数の事例に基づく頻度予測と個別の事例に基づく確率予測について検討を行う

Overview

- 実験 1a, 実験 1b
 - 被験者には、10 の課題 (学術的課題, 個人計画, 家事) を生成させた
 - 各課題に対して、課題完了日の見積もりを行わせた
 - 予測方法
 - ・ 確率予測
 - 各課題が各課題完了予測日までに完了する確率を記述させた
 - ・ 頻度予測
 - いくつの課題が課題完了予測日までに完了しているか見積もらせた
- 実験 1a
 - 被験者：夏期コースの授業を受けた大学生 71 人
 - 数ヶ月で完了する 10 の課題を授業中に考えさせ、各課題に対して、課題完了予測日、確率／頻度予測を記述させた
 - 夏期講習終了後、電話インタビューにより、課題完了日の確認を行った
- 実験 1b
 - 被験者：秋期コースの授業を受けた大学生 78 人
 - 1ヶ月で完了する 10 の課題を授業中に考えさせ、課題完了予測日、確率／頻度予測を記述させた
 - 1ヵ月後、実験者が与えた用紙に課題完了日を記述して提出させた

- 実験 1c
 - 被験者には，1つの課題を生成，遂行させた
 - 課題に対して，課題完了日の見積もりを行わせた
 - 予測方法
 - ・ 確率予測
 - 課題が課題完了予測日までに完了する確率を記述させた
 - ・ 頻度予測
 - 類似した課題が 10 ある場合，いくつの課題が課題完了予測日までに完了しているか見積もらせた
 - 被験者：冬季コースの授業を受けた大学生 139 人
 - ・ 2～4 週間で完了する課題 1つを授業中に考えさせ，課題に対して，課題完了予測日，確率／頻度予測を記述させた
 - ・ 4 週間後，電話インタビューにより，課題完了日の確認を行った

Results and Discussion

- Table 1 は，各実験の確率予測，頻度予測，予測通り完了した割合

TABLE 1
Mean Estimates of Task Completion and Actual Task Completion as a Function of Response Format and Question Order

Study/question order	Estimated percentage complete		Percentage complete
	Probability	Frequency	
Study 1a: Summer study			
Probability first	<u>72.9</u>	61.7	48.2
Frequency first	73.8	<u>68.3</u>	47.6
Study 1b: Autumn study			
Probability first	<u>73.0</u>	61.6	37.8
Frequency first	79.7	<u>73.3</u>	42.8
Study 1c: Winter (single-item) study			
Probability first	<u>73.8</u>	68.3	42.4
Frequency first	83.4	<u>74.3</u>	31.5

Note. The first two columns of numbers present the means of participants' estimates concerning the probability that each task (probability estimate) and how many of the tasks (frequency estimate) would be completed by the predicted time. Frequency estimates have been converted to percentages for comparability. The final column indicates the mean percentage of tasks actually completed by the predicted time. Underlined entries represent the first estimate made by participants.

- 全体的に、30%~40%の誤差で楽観的予測がなされた
- 予測通りに課題を完了させた割合
 - ・ 実験 1a
 - 確率予測：7%，頻度予測：24%
 - ・ 実験 1b
 - 確率予測：2%，頻度予測：12%
- 順序の効果がみられた
 - ・ 各実験ごとに被験者間でみた場合、1回目の確率予測と頻度予測の割合が近似
- 予測方法の効果がみられた
 - ・ 被験者内の確率予測と頻度予測を比較（実験 1a~1c の平均）
 - 1回目が確率予測の場合、確率予測(73.3%) > 頻度予測(64.6%)
 ($p < .001$ 両側 t 検定)
 - 1回目が頻度予測の場合、頻度予測(72.3%) < 確率予測(79.6%)
 ($p < .001$ 両側 t 検定)

- Figure 1 は、実験 1a と 1b の確率／頻度予測の平均を示した図

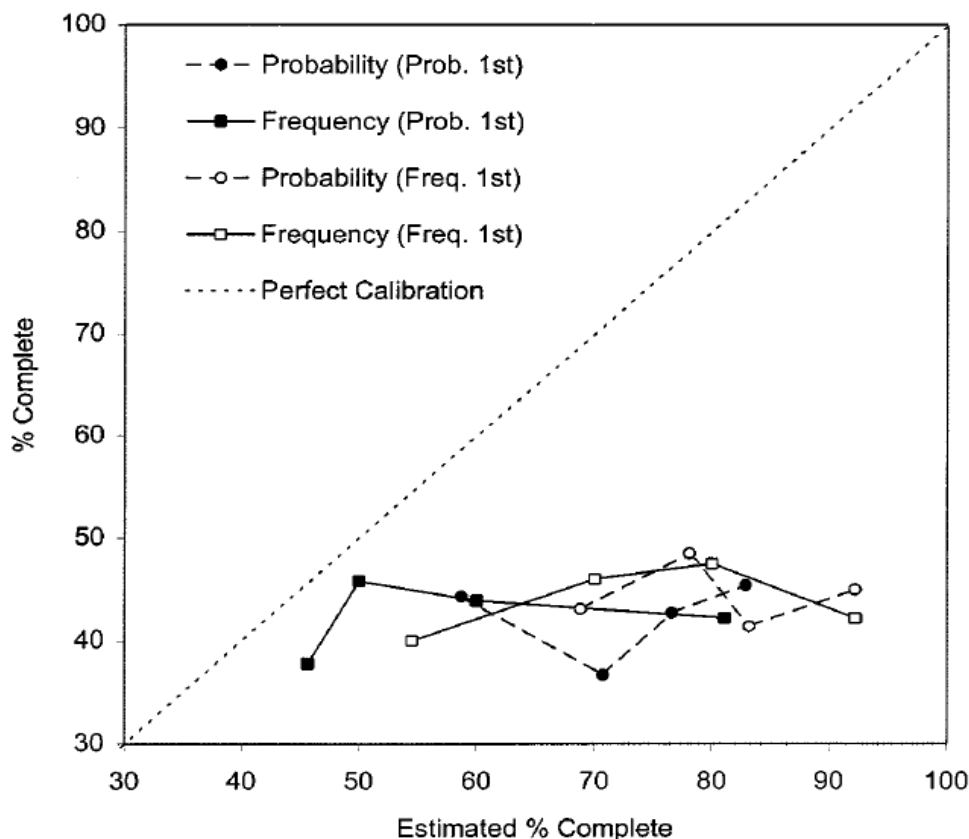


FIG. 1. Between-subject calibration curves for frequency estimates and mean probability estimates of task completion as a function of question order (Studies 1a and 1b combined). Frequency estimates have been converted into percentages. Each point on the curve represents the average response for approximately one-quarter of the participants within each condition.

- 4組のラインは、perfect calibration の下に位置した
 - ・ 確率予測，頻度予測は，共に楽観的に見積もられた
- 課題を完了した割合と，完了を予測した確率に相関関係はない
- 低い割合で課題の完了を予測したにも関わらず，高い割合で完了を予測したときよりも，実際には，高い割合で課題が完了されている場合がある

- Figure 2 は，実験 1a と 1b の確率予測の平均を示した図

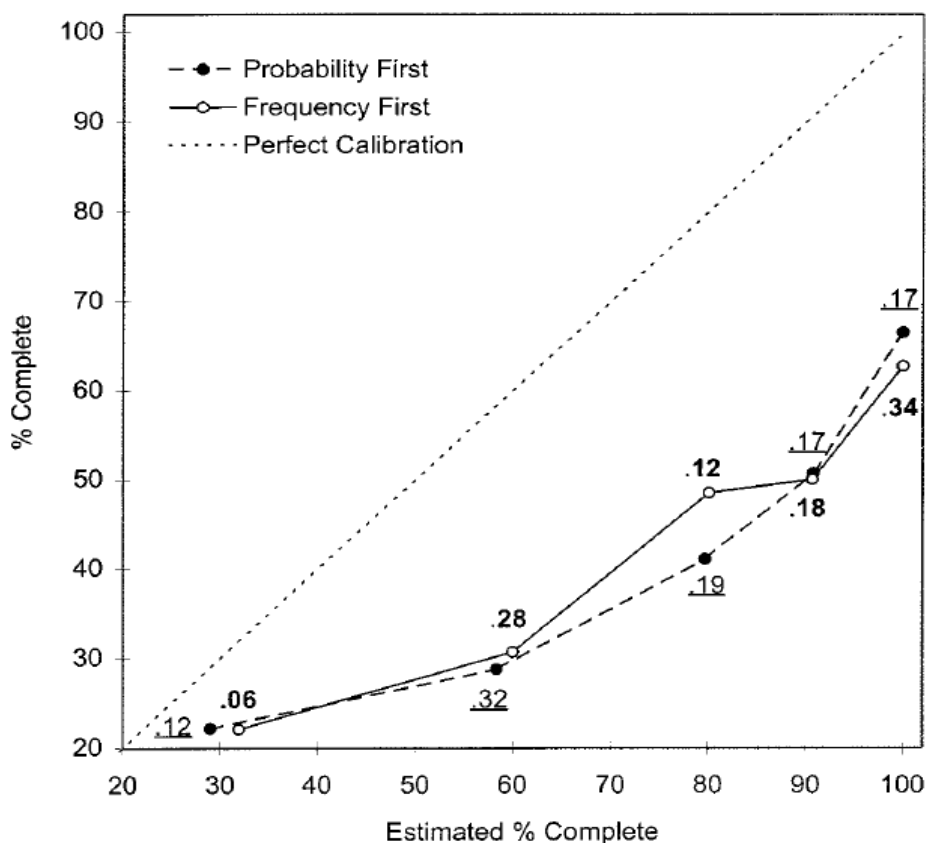


FIG. 2. Calibration curves for single-item probability estimates of task completion as a function of question order (Studies 1a and 1b combined). Numbers indicate the proportion of responses represented by each point on the curve within each question order, underlined numbers for the probability-first order and bold numbers for the frequency-first order.

- 2組のラインは，perfect calibration の下に位置したが，傾きがみられる
 - ・ 確率予測を行ったとき，完了しやすい，または，完了しにくいという予測能力がみられた (kim's d: 確率予測 first (.38)，頻度予測 first (.31))

- Table 2 は，実験 1c において，確率／頻度予測を行った際の発話をもとに，グループ分けしたもの

TABLE 2
 Percentage of Respondents Using Each Thought Category for
 Probability and Frequency Estimates of Task Completion (Study
 1c: Winter Study)

Thought categories	Probability	Frequency
Inside perspective		
Task requirements	25	32
Planning	27	17
Hopes	22	6
Self-efficacy	9	6
Hedonic considerations	7	6
Outside perspective		
Past experiences	39	44
Personal disposition	3	13
Other		
Relative priority of tasks	24	24
Competing tasks	29	24
Deadline	17	9

- Hopes に関する発話が、有意に確率予測を行う際、多かった($p < .02$ χ^2 検定)
- わずかな差はいくつかのカテゴリーで見られるが、確率予測と頻度予測の間に明確な差はみられなかった
- まとめ
 - これらの結果からは、頻度予測が、確率予測よりも優れているとはいえない
 - 日常的な場面では、予測を行う際の統計的性質を隠してしまう可能性がある
 - ・ 以下の実験は、実験室での実験を行う
 - 実験 1 では、様々な課題が用いられた。課題の複雑性が予測に影響した可能性がある
 - ・ 課題の複雑性を調整を行う
- Confidence in knowledge
 - Gigerenzer (1991)
 - ・ overconfidence は、確率予測ではみられるが、頻度予測ではみられない
 - Griffin & Tversky (1992), Lichten, Fischhoff & Phillips (1982), Suantak, Bolger & Ferrell (1996)
 - ・ overconfidence は、判断材料が少なく、問題の難易度が高いときにみられる
 - ・ underconfidence は、判断材料が多く、問題の難易度が低いときにみられる
 - Griffin & Tversky (1992)

- confidence-frequency discrepancy
 - 課題として、アメリカの州を選択する地理の問題を使用し、頻度予測と実際の正解数との相違を確認した
 - Griffin & Tversky (1992) は、課題の複雑性を調整しなかった
- 実験 2 では、課題の複雑性を調整を行う

Experiment 2

Overview

- 被験者には、2 択の一般的な問題を 20 問解答させた
(Oregon Decision Research Group により作成された問題)
- 各課題に対して、信頼度の記述と、課題後に、課題全体の正解率を見積もらせた
- 実験 2a
 - 被験者：大学生 221 人
 - 課題の難易度は、高いものから低いものまで、調整された 20 問の問題を使用
 - 予測方法
 - 頻度予測
 - 課題後に、何問正解したか見積もらせた
 - 更に、どのようにその見積もりを行ったのか説明を行わせた
 - 確率予測
 - 何%正解であったか見積もらせた
 - 各問題に関して、何%の確率で正解であったか見積もらせた
- 実験 2b
 - 被験者：大学生 130 人
 - 難易度の低い問題、また、実験 2a で正確さと予測の割合が高かった問題は、難易度の高い問題に取り換えた 20 問を使用
 - 他は実験 2a と同様
- 実験 2c
 - 被験者：大学生 71 人
 - 難易度の高い 10 問、低い 10 問題の 20 問を使用
 - 予測方法
 - 頻度確率 first
 - 最初の 10 問(難易度が高い)を解答後、10 問中、何問正解するか見積もらせた
 - 後の 10 問(難易度が低い)を解答後、後の 10 問中、何問正解したか見積もら

- せた
- 各問題に関して、何%の確率で正解であったか見積もらせた
 - 確率予測 first
 - 各問題の解答後、何%の確率で正解であったか見積もらせた
 - 全ての問題を解答後、最初の 10 問について何問正解したか、後の 10 問について何問正解したか見積もらせた

Results and Discussion

- Table 3 は、各実験の確率予測、頻度予測、正解した割合

TABLE 3
 Mean Probability Estimates, Frequency Estimates, and Correct Answers for
 General-Knowledge Questions

	Estimated probability	Estimated frequency	Percentage correct
Study 2a: Moderate	67.2	60.2	66.5
Study 2b: Difficult	64.7	59.0	56.1
Study 2c			
Moderate Set	74.6	65.0	75.7
Difficult Set	64.2	55.9	51.5

Note. The first two columns of numbers present the means of participants' estimates concerning the probability that each answer was correct (item-specific probability estimate) and how many of the answers were correct (aggregate frequency estimate). Frequency estimates have been converted to percentages for comparability. The final column indicates the mean percentage of correct answers.

- 実験 2a では、頻度予測は、有意に underconfident であった ($p < .001$ 両側 t 検定)
- 実験 2b では (課題の複雑性が高い場合)、確率予測 ($p < .001$ 両側 t 検定)、頻度予測 ($p < .02$ 両側 t 検定) は有意に overconfident であった
- 実験 2c で、課題の複雑性は調整されたが、確率予測、頻度予測でその差はあらわれなかった
 - 実際に正解した割合の誤差は 24.2%
 - 確率予測では 10.4%の誤差、頻度予測では 10.9% (9.1%) の誤差
- 実験 2a
 - 課題の複雑性が低い場合

- 頻度予測よりも、確率予測の方が正確に予測した ($p < .001$ 両側 t 検定)
- 課題の複雑性が高い場合
 - 逆に、頻度予測の方が正確に予測した ($p < .001$ 両側 t 検定)
- 確率予測、頻度予測の両方で妥当な判断はなされなかった

- Figure 3 は、実験 2a と 2b の確率／頻度予測の平均を示した図

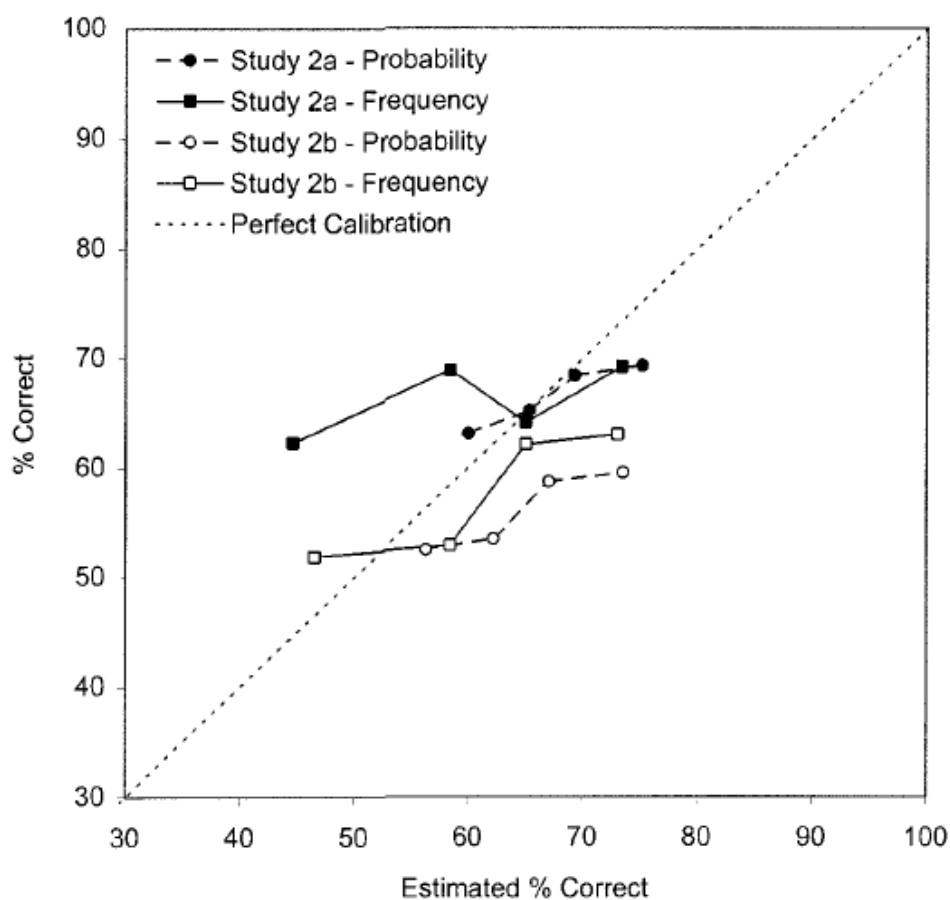


FIG. 3. Between-subject calibration curves for frequency estimates and mean probability estimates for general knowledge questions in Study 2a (moderately difficult items) and Study 2b (difficult items). Frequency estimates have been converted into percentages. Each point on a curve represents the average for approximately one-quarter of the participants within each study.

- 実験 2a の方が難易度は低かったが、実験 2b の方が成績を下回った
- 頻度予測は、左へ移行した
 - 確率予測よりも、頻度予測で、低い見積もりが行われた

- Figure 4 は、実験 2c の確率／頻度予測の平均を示した図

- 課題の複雑性の影響を大きく受けている
- 頻度予測は、左へ移行した
 - ・ 確率予測よりも、頻度予測で、低い見積もりが行われた

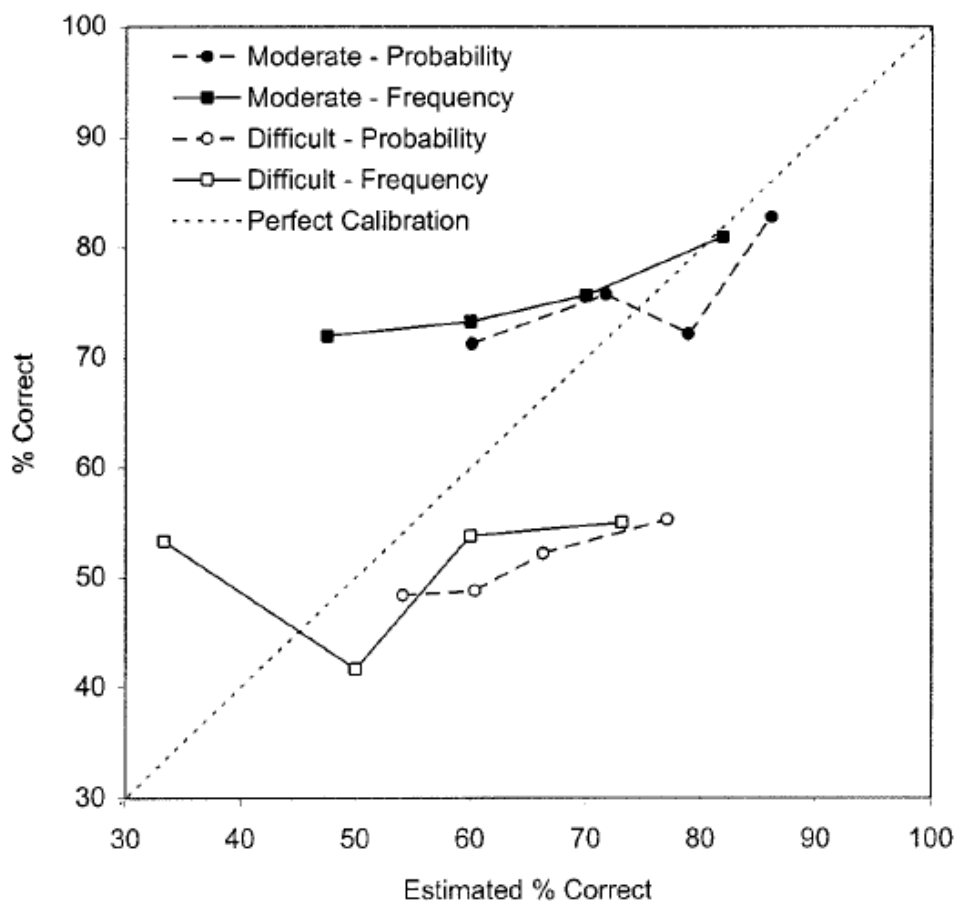


FIG. 4. Between-subject calibration curves for frequency estimates and mean probability estimates for general knowledge questions in Study 2c, with separate curves for the moderately easy and the difficult item sets. Frequency estimates have been converted into percentages. Each point on a curve represents the average for approximately one-quarter of the participants.

- この結果から、確率予測よりも、頻度予測を行う方が、予測される値は低くなることが確認された
- Table 4 は、実験 2a と 2b において、確率／頻度予測についての説明をもとに、予測に用いた方略を示したもの
 - Counting：正解であると予測した問題数を数える
 - ・ Counting を行った被験者は、行わなかった被験者よりも、頻度予測を低く見積もり、実際の正解率と予測とのかい離が大きい
 (実験 2a, $p < .001$ 両側 t 検定; 実験 2b $p < .001$ 両側 t 検定)

TABLE 4
Mean Probability Estimates, Frequency Estimates, and Correct Answers for General-Knowledge Questions by Use of Thought Categories (Studies 2a and 2b)

Study/thought categories	<i>n</i>	Estimated probability	Estimated frequency	Percentage correct
Study 2a: Moderate items				
Counting				
No	176	67.9	62.5	67.0
Yes	45	64.6	51.6	64.4
Integration with chance				
No	76	66.3	54.3	65.3
Yes	145	67.7	63.3	67.1
Background				
No	154	66.9	61.1	65.9
Yes	67	67.9	58.3	67.7
Study 2b: Difficult items				
Counting				
No	91	65.9	62.3	56.9
Yes	39	62.0	51.4	54.2
Integration with chance				
No	56	62.5	53.7	54.1
Yes	74	66.4	63.0	57.6
Background				
No	98	65.4	59.6	57.3
Yes	32	62.6	57.2	52.3

Notes. Rows labeled No represent participants who did not use the thought category and rows labeled Yes represent participants who did. The first two columns of means represent participants' estimates concerning the probability that each answer was correct (item-specific probability estimate) and how many of the answers were correct (aggregate frequency estimate). Frequency estimates have been converted to percentages for comparability. The final column presents the mean percentage of correct answers.

- Integration with chance : Counting に加え，正解数の推測を行う
 - Integration を行った被験者は，行わなかった被験者よりも，頻度予測を高く見積もり，実際の正解率と予測とのかい離が小さい
(実験 2a, $p < .001$ 両側 t 検定; 実験 2b $p < .001$ 両側 t 検定)
- Background : 課題の複雑性や，自らの能力などを参照する
 - Background を行った被験者と，行わなかった被験者に有意な差はみられなかった
- まとめ
 - 頻度予測と実際の正解との相違は，部分的な知識を無視することで生じる

- ・ 今回の実験では、正解と予測した問題数を数え、更に正解数の推測を行うことで、頻度予測と実際の正解との相違をなくすことに成功した
- **Background** として、課題の複雑性について言及しても、頻度予測と実際の正解との相違に影響はしなかった
- 頻度予測は低く見積もられ、確率予測より高く見積もられることはなかった
 - ・ 頻度予測はバイアスの影響を受けないとはいえない
 - ・ **overconfidence** を生む確率予測と **underconfidence** を生む頻度予測の調整が必要であろう
- **Frequency and base rate use**
 - 分類的な予測は、一般的に用いられる
 - 分類的な予測は、特定の情報と一般性、珍しさに関する基準率の情報を組み合わせて判断される
 - 実験 3 では、基準率における確率予想と頻度予想の効果を検討する
 - **Kahneman & Tversky (1973)**
 - ・ **Lawyer-engineer paradigm**
 - 弁護士 70 人とエンジニア 30 人の状況が設定される
 - 被験者には、個人情報を与えられ、その人物が弁護士か、エンジニアかを推測する
 - 被験者は、基準率を無視し、個人情報による判断を行う

Experiment 3

Overview

- 課題
 - 透明の箱には、100 個のボール（70 個は白、30 個は緑）が入れられている
 - 白のボールに弁護士の個人情報、緑のボールにエンジニアの個人情報が入っていると告げられる
 - ・ カウンターバランスをとって、白がエンジニア、緑が弁護士の場合も設定
 - ・ 個人情報は全て一人のエンジニア (**Jack**) に関する情報である
 - 被験者は、箱の中身を見て、まずボールが何個ずつあるか推測する
 - 被験者は、目隠しをして、ボールを取り出し、情報を読む
 - ・ ボールの色は知らされない
- 実験 3a, 3b
 - 予測方法
 - ・ 確率予測

- 読まれた情報の人物がエンジニアである確率を答える
- 頻度予測
 - 10個ボールを取り出した場合、同様のエンジニアに関する情報が、いくつ得られるかを答える
- 課題の後に、被験者が思い描く弁護士、エンジニアと読まれた情報の一致度を1~11(1: 弁護士~11: エンジニア)で記述する
- 実験 3c
 - 予測方法
 - 確率/頻度予測のいずれかの予測を行う
 - 課題の後に、被験者が思い描く弁護士、エンジニアと読まれた情報の一致度を1~10(1: 弁護士~10: エンジニア)で記述する

Results and Discussion

- Figure 5は、実験 3a, 3b, 3cの確率/頻度予測の中央値を示した図

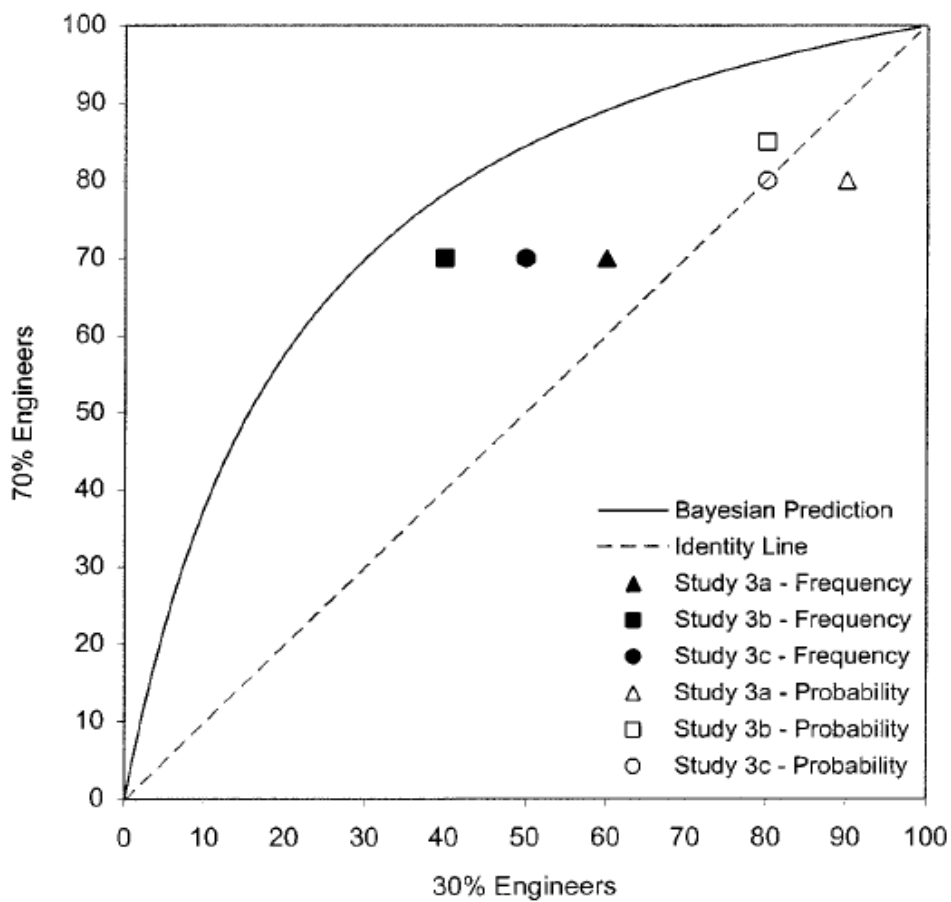


FIG. 5. Median probability and frequency estimates (engineer) as a function of base rates (30% Engineers vs. 70% Engineers) for Studies 3a, 3b, and 3c. The identity line indicates complete neglect of base rates. The curved line indicates base rates used according to Bayes' rule.

- 確率予測は，基準率を無視した結果を示した
- 基準率は，確率予測よりも，頻度予測に影響した
 - ・ しかし，ベイズによる予測に従う結果はなかった
- 基準率に関わらず，頻度予測は，確率予測よりも，低い値を示した
 ($p < .001$ 両側 t 検定)

● 頻度予測と基準率

$$\log \left[\frac{P(H_E|D)}{P(H_L|D)} \right] = \log \left[\frac{P(D|H_E)}{P(D|H_L)} \right] + \log \left[\frac{P(E)}{P(L)} \right],$$

エンジニアである確率

情報の一致度

基準率

TABLE 5
 Regression Weights Indicating the Relative Use of Base Rate and Case-Specific Information for Probability and Frequency Estimates

Study/response format	Base rate	Case specific
Study 3a		
Probability	.03	.93
Frequency	.30	.47
Study 3b		
Probability	.19	.99
Frequency	.62	.39
Study 3c		
Probability	.07	.99
Frequency	.48	.44

- Table 5 は，確率／頻度予測における回帰係数を示す
 - 確率予測を行うと，個別情報に重きを置く
 - 頻度予測を行うと，個別情報の使用が減少する
- Table 6 は，実験 3b と 3c における確率／頻度予測の結果
 - 確率予測を行った場合，多くの被験者は，個別情報のみに言及した
 - 頻度予測を行った場合，個別情報と基準率に言及する被験者の割合は同程度

TABLE 6

Percentage of Respondents Using Each Information Category and Median Probability or Frequency Estimates as a Function of Response Format, Thought Categories, and Base Rate Condition

Response format/ information categories	% of Respondents	Median probability or frequency estimate	
		70 engineers	30 engineers
Study 3b: Laboratory demonstration			
Probability			
Base rate	15	63	40
Case specific	78	89	80
Both	6	75	—
Frequency			
Base rate	48	70	30
Case specific	43	75	70
Both	7	75	—
Study 3c: Laboratory questionnaire			
Probability			
Base rate	15	65	30
Case specific	81	80	80
Both	4	70	—
Frequency			
Base rate	48	70	30
Case specific	32	80	70
Both	14	80	50

Note. Table entries are the medians for participants' estimates of how likely it is that the description is of an engineer (probability estimate) or how many similar descriptions would be engineers (frequency estimate). Frequency estimates have been converted to percentages for comparison with probability estimates. Dashes indicate that no participants in the condition used the information category.

General Discussion

- 情報の集約だけでは、的確な予測はなされない
 - しかし、様々な単純な方略が有益な情報を与える
- 今回の実験結果は、これまでに示されてきた結果とは異なる
 - 頻度予測も認知的錯覚 (cognitive illusion) の影響を受ける
- 判断についての研究は、判断される確率、価値、嗜好についてではなく、判断を行う文脈を計画することにある。
 - 様々な要因が、規範的に適切／不適切なプロセスの使用に影響する。様々な方法により検討を行い、特定の明確な状況を提示することが判断の理解につながる