Analytic Confidence and Political Decision Making

Supporting Information

Contents

1. Full text of vignettes and experimental manipulations

Below we present the full text of each scenario, with experimental manipulations embedded. In addition, we randomized the ordering of the three conceptions of confidence.

Hostage rescue scenario

> The U.S. military is searching for five American citizens held hostage by a rebel group overseas. They receive information suggesting that the hostages are being held in a rural compound. Analysts can tell that the compound is being used by the rebel group, but they have difficulty confirming that the hostages are present. Special forces officers expect that a raid on the compound will meet armed resistance. They do not want to put their soldiers in harm's way if the hostages are not present, but if they delay action too long, the rebels might move the hostages to a different location.
>
> After careful deliberation, a group of intelligence analysts assesses that there is a [60 / 75] percent chance that the hostages are being held inside this compound. The analysts explain that they have [little / a large amount of] reliable evidence on which to base their judgment; that there is [minimal / significant] disagreement among them about the chances that the hostages are present; and that they [do not believe their assessment would substantially change / believe their assessment could substantially change] if they continue to investigate the compound.

Drone strike scenario

> U.S. intelligence officials are attempting to locate a high-ranking terrorist. Drone operators say that they have found a man who meets their target's description. He is driving alone, in a deserted area. However, it is always difficult to confirm a target's identity using remote surveillance. U.S. officials worry that the man could be an innocent civilian, and analysts cannot rule out this possibility. At the

same time, if the target is indeed a high-ranking terrorist, then delaying a strike could give him a chance to escape.

After careful deliberation, a group of intelligence analysts assesses that there is a [80 / 95] percent chance that this man is a high-ranking terrorist. The analysts explain that they have [little / a large amount of] reliable evidence on which to base their judgment; that there is [minimal / significant] disagreement among them about the chances that the target is a civilian; and that they [do not believe their assessment would substantially change / believe their assessment could substantially change] if they continue to track this suspect.

Terrorism scenario

U.S. intelligence analysts receive information about a potential terrorist attack. Informants warn that terrorists plan to use a new form of explosive that is extremely difficult to detect against several flights departing from California. They say that the plotters are already inside the United States and that they could strike at any time. Yet analysts have reasons to doubt that the plot is real. In particular, terrorists may be planting false information to trick the U.S. government into restricting air travel, which would cause panic and economic damage.

After careful deliberation, a group of intelligence analysts assesses that there is a [10 / 25] percent chance that the plot is real. The analysts explain that they have [little / a large amount of] reliable evidence on which to base their judgment; that there is [minimal / significant] disagreement among them about the chances that the plot is real; and that they [do not believe their assessment would substantially change / believe their assessment could substantially change] if they continue to investigate the plot.

2. Distribution of responses to scenarios

Figures S1a-S1c present the distribution of responses for the four evaluations we elicited on each scenario.

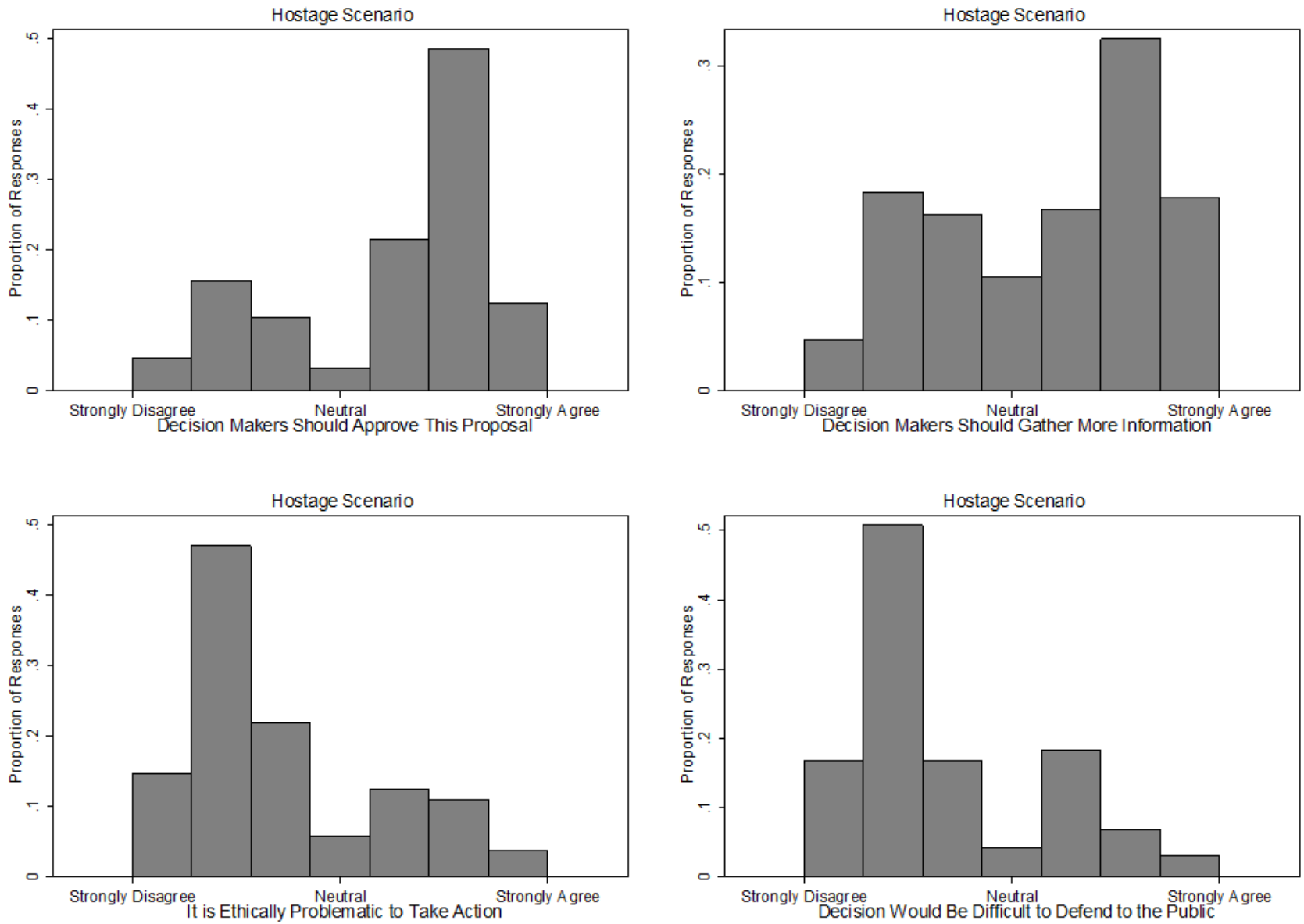


Figure S1a. Responses to hostage rescue scenario

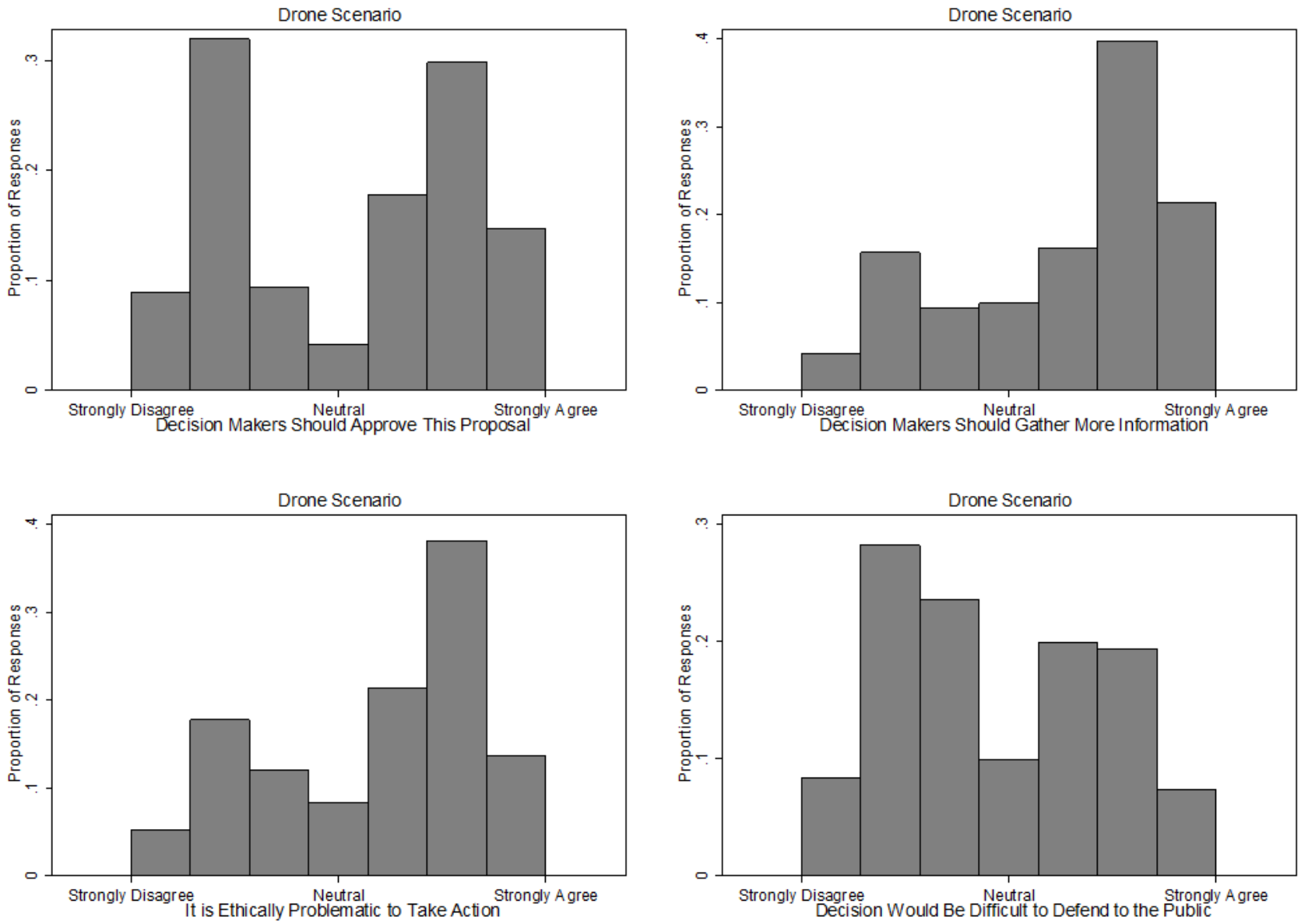


Figure S1b. Responses to drone strike scenario

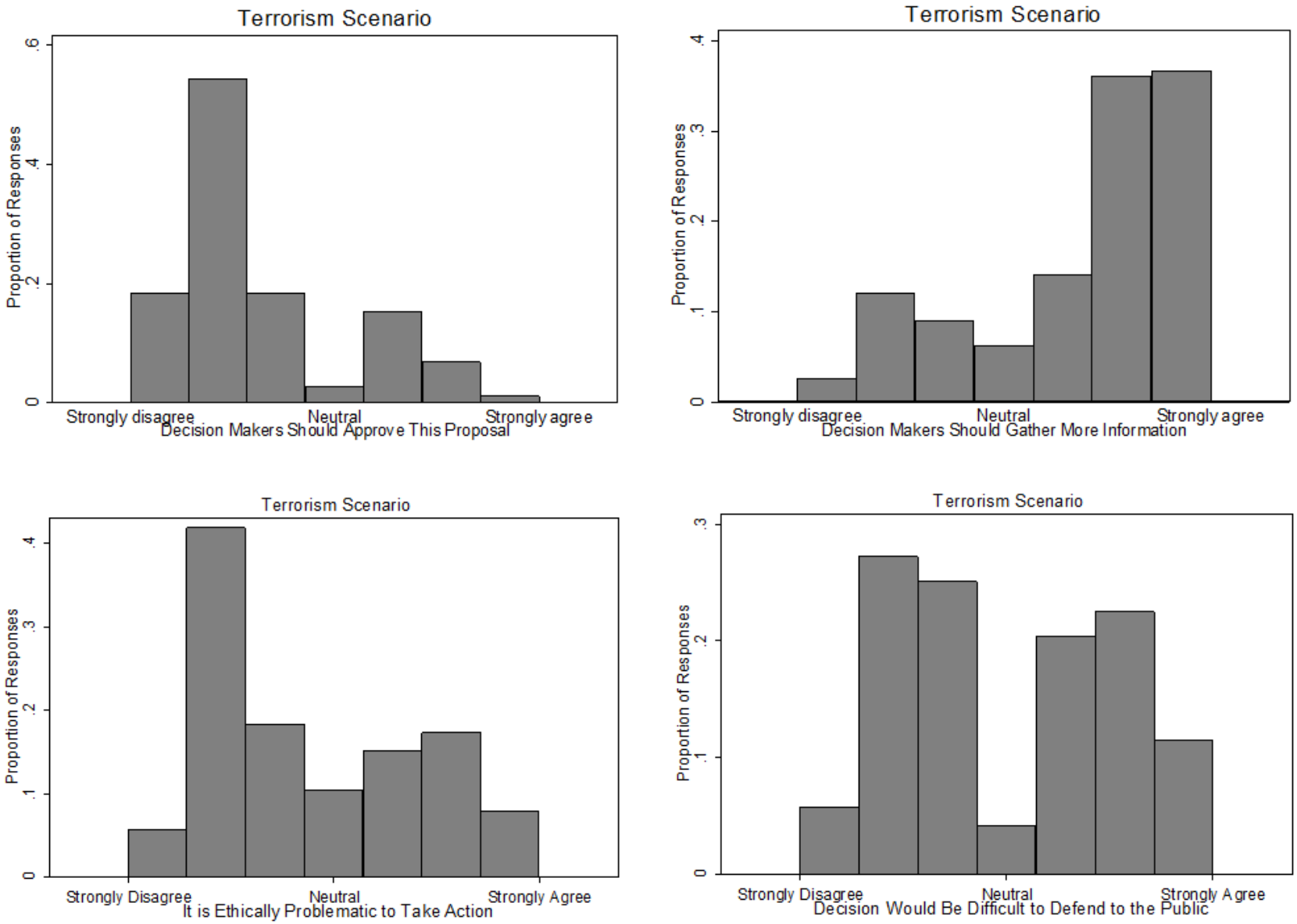


Figure S1c. Responses to terrorism scenario

3. Full results for survey experiment 2

The main text of the paper summarized ordinary least squares regressions predicting responses to our first survey experiment. Here are the full results for that analysis. We include respondent fixed effects (not shown) and cluster standard errors by respondent.

	Model 1 DV: Support for taking action	*Model 2* DV: Willingness to delay action	*Model 3* DV: Decision is ethically problematic	*Model 4* DV: Decision is politically problematic
Probability	0.82 (.13)***	-0.53 (.13)***	-0.38 (.13)**	-0.51 (.13)***
Reliability	0.25 (.13)	-0.28 (.13)*	-0.28 (.14)*	-0.46 (.13)***
Disagreement	0.39 (.13)**	-0.46 (.13)***	-0.30 (.13)*	-0.38 (.12)**
Responsiveness	0.27 (.13)*	-0.59 (.15)***	-0.05 (.14)	-0.14 (.13)
Hostage	2.19 (.15)***	-0.79 (.15)***	-0.60 (.13)***	-1.14 (.14)***
Drone	1.53 (.17)***	-0.48 (.16)**	1.01 (.17)***	-0.28 (.16)
Constant	1.79 (.16)***	6.31 (.18)***	4.13 (.18)***	4.82 (.18)***
N	669	669	669	669
R^2 (overall)	0.27	0.09	0.15	0.12

Table S1 presents ordinary least squares regressions predicting reactions to national security decisions. All dependent variables are measured on 7-point scales. Respondent fixed effects not shown. $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Standard errors clustered by respondent.

Table S1. Responses to national security scenarios (OLS)

4. Alternative specifications for survey experiment 1

The following tables present alternative specifications for survey experiment 1. We begin by estimating treatment effects using ordered logit. Compared to the results shown in Table S1, the only two substantial changes are that the coefficient for *Responsiveness* loses statistical significance in Model 1 (p=0.09), and the coefficient for *Reliability* loses statistical significance in Model 2 (p=0.06), but neither of these changes would influence our main conclusions, and in both cases p-values still fall close to standard significance thresholds.

	Model 1	*Model 2*	*Model 3*	*Model 4*
	DV: Support for taking action	DV: Willingness to delay action	DV: Decision is ethically problematic	DV: Decision is politically problematic
Probability	0.94 (.15)***	-0.67 (.16)***	-0.46 (.15)**	-0.62 (.15)***
Reliability	0.25 (.14)	-0.32 (.15)*	-0.29 (.16)	-0.54 (.15)***
Disagreement	0.46 (.13)***	-0.49 (.17)**	-0.35 (.15)*	-0.47 (.14)***
Responsiveness	0.24 (.14)	-0.80 (.17)***	-0.08 (.16)	-0.15 (.16)
Hostage	2.20 (.16)***	-1.03 (.18)***	-0.73 (.16)***	-1.41 (.17)***
Drone	1.55 (.20)***	-0.73 (.20)***	1.07 (.20)***	-0.31 (.18)
Cut points	-0.54, 1.59, 2.16, 2.31, 3.14, 5.10	-5.80, -3.82, -3.11, -2.65, -1.92, -0.14	-3.58, -1.16, -0.37, 0.00, 0.83, 2.63	-4.36, -2.03, -1.06, -0.78, 0.25, 1.77
N	669	669	669	669

Table S2 presents ordered logit models predicting reactions to national security decisions. All dependent variables are measured on 7-point scales. Respondent fixed effects not shown. * p<0.05, **p<0.01, ***p<0.001. Standard errors clustered by respondent.

Table S2. Responses to national security scenarios (ordered logit)

Next, we estimate experimental treatments using simple two-way t-tests. This is a less credible way to estimate quantities of interest: since our central claim is that respondents will process conceptions of confidence simultaneously, we believe that the most appropriate way to test this hypothesis is to use multivariate regression. As all experimental treatments were randomized, however, we can plausibly examine these results in isolation. As with Table S2, these results sustain the same conclusions presented in the paper.

	Model 1	*Model 2*	*Model 3*	*Model 4*
	DV: Support for taking action	DV: Willingness to delay action	DV: Decision is ethically problematic	DV: Decision is politically problematic
Reliability	0.27 (p=0.08)	-0.28 (p=0.05)	-0.28 (p=0.06)	-0.46 (p=0.001)
Disagreement	0.38 (p=0.01)	-0.44 (p=0.002)	-0.33 (p=0.02)	-0.43 (p=0.002)
Responsiveness	0.27 (p=0.08)	-0.61 (p<0.001)	-0.05 (p=0.71)	-0.15 (p=0.29)

Table S3 presents treatment effects estimated using two-way t-tests, describing how varying each conception of confidence changed response measures, on average.

Table S3. Responses to national security scenarios (two-way t-tests)

Table S4 replicates our previous multivariate findings without respondent fixed effects. All coefficients and standard errors are substantively similar, though this causes the *p*-values for *Reliability* (p=0.041 to p=0.053) and *Disagreement* (p=0.027 to p=0.064) to fall just over the standard threshold for statistical significance in Model 3.

	Model 1 DV: Support for taking action	*Model 2* DV: Willingness to delay action	*Model 3* DV: Decision is ethically problematic	*Model 4* DV: Decision is politically problematic
Probability	0.82 (.13)***	-0.50 (.14)***	-0.34 (.13)*	-0.53 (.13)***
Reliability	0.25 (.13)	-0.28 (.14)*	-0.27 (.14)	-0.46 (.13)***
Disagreement	0.38 (.13)**	-0.43 (.14)**	-0.26 (.14)	-0.41 (.13)**
Responsiveness	0.27 (.13)*	-0.60 (.15)***	-0.03 (.14)	-0.14 (.13)
Hostage	2.20 (.15)***	-0.79 (.15)***	-0.60 (.13)***	-1.14 (.14)***
Drone	1.53 (.17)***	-0.48 (.16)**	1.01 (.17)***	-0.29 (.16)
Constant	1.79 (.16)***	6.29 (.18)***	4.08 (.19)***	4.83 (.18)***
N	669	669	669	669
R^2 (overall)	0.27	0.09	0.15	0.12

Table S4 presents ordinary least squares regressions predicting reactions to national security decisions. All dependent variables are measured on 7-point scales. $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Standard errors clustered by respondent.

Table S4. Responses to national security scenarios (OLS, no respondent fixed effects)

5. Complete wording of questions for survey experiment 2

Below we present the wording and formatting of each question as they appeared to respondents in our survey. There are 20 questions in total. Each respondents answered a randomly-selected subset of 10 questions, presented in random order.

We also posed an attention check question asking what are the chances that two plus two equals four, and instructing respondents to answer 100 percent. Ninety-seven percent of respondents passed this attention check. Our results are robust to excluding the remaining respondents from the data set.

1. In your opinion, what are the chances that **Hillary Clinton will win the 2016 presidential election?** A recent RealClearPolitics poll puts Clinton five points ahead of Donald Trump, but polls have previously underestimated Trump's performance.

2. In your opinion, what are the chances that **the U.S. Senate will approve President Obama's nomination of Merrick Garland to the U.S. Supreme Court by the end of 2016?** Senate Majority Leader Mitch McConnell has vowed not to hold a vote on this issue, but some Republican Senators have said that they are willing to consider his nomination.

3. In your opinion, what are the chances that **President Barack Obama's approval rating will be above 50% at the end of 2016?** Obama's approval rating is currently at 51% according to Gallup.

4. In your opinion, what are the chances that **the U.S. unemployment rate will be below 5% at the end of 2016?** The U.S. unemployment rate is roughly 5.5% today, down from a high of roughly 10% in 2009.

5. In your opinion, what are the chances that **a federal court will restore the Washington Redskins trademark by the end of 2016?** The U.S. Patent Office canceled the Redskins' trademark on the grounds that the name is offensive. The football team appealed this ruling, arguing that many other controversial brand names currently receive trademark protection.

6. In your opinion, what are the chances that **a third party candidate will win more than 10% of the popular vote in the 2016 presidential election?** The last third-party presidential candidate to achieve this feat was Ross Perot, who won 19% of the popular vote in 1992.

7. In your opinion, what are the chances that **the U.S. Congress will pass any restriction on access to firearms by the end of 2016?** Pressure has been building for such measures in recent years, but the Senate recently voted against several gun control proposals.

8. In your opinion, what are the chances that **the average price of a gallon of regular, unleaded gasoline in the United States will still be below $2.50 at the end of 2016?** That price is currently about $2.30.

9. In your opinion, what are the chances that **a military court will sentence U.S. Army Sergeant Bowe Bergdahl to more than five years of imprisonment?** Bergdahl is currently facing courtmartial for leaving his post in Afghanistan and he is subject to a potential life sentence.

10. In your opinion, what are the chances that **the global average temperature will be warmer in 2016 than it was in 2015?** 2015 was the hottest year on record, but climate skeptics say this was an exception rather than a trend.

11. In your opinion, what are the chances that **Britain will formally exit the European Union by the end of 2016?** Britain's voters approved a referendum in favor of leaving the EU, but that referendum is nonbinding and the country has no explicit timetable for carrying it out.

12. In your opinion, what are the chances that **Bashar al-Assad will no longer be Syria's president by the end of 2016?** Rebels have sought to oust al-Assad for the past four years, and President Obama has called for him to step down, but the war has recently come to a standstill.

13. In your opinion, what are the chances that **the United States will accept more than 7,500 Syrian refugees by the end of 2016?** President Obama committed to accepting 10,000 Syrian refugees this year, but his administration has only taken about 2,500 Syrian refugees so far, and many voters oppose admitting more.

14. In your opinion, what are the chances that **the United States will close the Guantanamo Bay prison camp by the end of 2016?** President Obama has repeatedly promised to close the prison, but 79 detainees currently remain at the site.

15. In your opinion, what are the chances that **the next United Nations Secretary General (who will be elected in 2016) will be a woman?** The United Nations has never had a female Secretary General and several member states are pushing to consider female nominees in this election.

16. In your opinion, what are the chances that **a terrorist event will kill more than 100 people in the United States by the end of 2016?** No terrorist attack has killed this many people in the United States since 9/11. The Orlando nightclub shooting in June killed 49.

17. In your opinion, what are the chances that **Russia's economy will shrink in 2016?** Russia's economy contracted by 5% in 2015 as a result of falling oil prices and international sanctions for the country's actions in Ukraine.

18. In your opinion, what are the chances that **U.S. forces will capture or kill the current head of Al Qaeda, Ayman al-Zawahiri, by the end of 2016?** The United States killed Al Qaeda's previous leader, Osama bin Laden, in 2011.

19. In your opinion, what are the chances that **Edward Snowden will return to the United States by the end of 2016?** In October 2015, Snowden stated that he would plead guilty to leaking classified documents if the United States government offered him a limited prison sentence.

20. In your opinion, what are the chances that **more than 10 U.S. soldiers will be killed in Iraq in 2016?** The United States currently has roughly 3,500 soldiers stationed in Iraq. Their current mission does not include direct combat, but some critics are pushing the White House to expand that mission.

6. Distribution of probability assessments from survey experiment 2

Figure S2 presents a histogram of probability assessments that respondents provided for survey experiment 2. Respondents generally provided probability estimates in intervals of ten or five percentage points, but unlike when eliciting "feeling thermometers" in other areas of survey research, we see no indication that responses cluster at two or three areas of the spectrum.

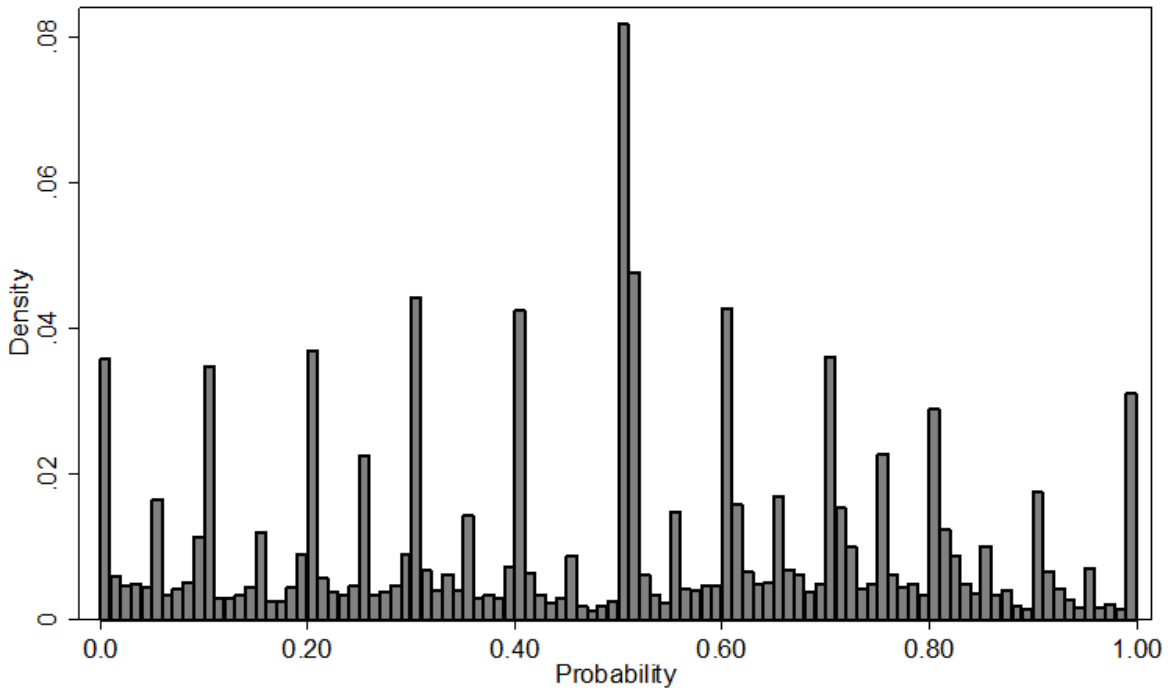


Figure S2. Distribution of probability assessments provided by respondents

Note that, since our method for eliciting probability assessments involved using sliders, respondents often gave values "near" round numbers, and it is likely that some of these responses suggest minor operator error instead of an attempt to offer highly-granular probabilities.

7. Additional descriptive statistics from survey experiment 2

In the paper, we briefly presented descriptive statistics indicating the extent to which confidence assessments varied within and across questions. Here, we present those data in more detail.

Range of reasonable opinion

Reliability of available evidence

	1	2	3	4	5	
1	4.6%	2.3%	0.8%	0.3%	0.3%	
2	4.8%	8.6%	1.8%	1.1%	0.1%	
3	3.6%	8.3%	7.0%	0.9%	0.1%	
4	7.8%	7.8%	18.9%	5.8%	0.7%	
5	4.1%	4.2%	1.6%	4.5%	2.9%	Correl: 0.27, *N*=10,000

Responsiveness to new information

Reliability of available evidence

	1	2	3	4	5	
1	5.4%	1.5%	0.6%	0.4%	0.5%	
2	6.9%	6.6%	1.4%	1.3%	0.1%	
3	4.7%	7.3%	6.4%	1.2%	0.2%	
4	8.8%	16.3%	5.0%	6.9%	1.2%	
5	3.2%	3.4%	1.6%	4.6%	4.3%	Correl: 0.34, *N*=10,000

Responsiveness to new information

Range of reasonable opinion

	1	2	3	4	5	
1	15.0%	6.0%	1.5%	1.5%	0.8%	
2	11.0%	12.0%	4.7%	4.4%	1.0%	
3	2.1%	5.0%	6.9%	2.2%	0.4%	
4	0.9%	2.9%	1.8%	5.7%	1.5%	
5	0.1%	0.3%	0.2%	0.8%	2.7%	Correl: 0.53, *N*=10,000

Figure S3. Correlations between pairs of attributes

	Probability	*Reliability of available evidence*	*Range of reasonable opinion*	*Responsiveness to new information*	*N*
1. Hillary Clinton elected President	64.41	3.87	2.24	2.88	517
2. Garland confirmed to Supreme Court	41.88	3.31	2.35	2.39	491
3. Obama approval rating over 50%	57.26	3.63	2.15	2.55	485
4. U.S. unemployment under 5%	40.53	3.56	2.25	2.24	528
5. Washington Redskins trademark restored	40.63	3.16	2.26	2.37	497
6. Third-party nominee wins 10% popular vote	38.98	3.64	2.44	2.54	487
7. Congress passes gun control measure	34.24	3.85	2.47	2.71	512
8. Avg. gasoline price under $2.50/gallon	58.98	3.48	2.18	2.08	495
9. Bergdahl sentenced to more than five years	56.31	3.00	2.14	1.98	498
10. Avg. global temp. higher in 2016 than in 2015	66.92	4.03	2.77	2.72	533
11. Britain formally exits European Union	59.62	3.82	2.42	2.50	484
12. Assad no longer Syria's president	40.52	2.98	2.15	2.03	516
13. U.S. takes more than 7,500 Syrian refugees	49.83	3.35	2.18	2.16	491
14. U.S. closes Guantanamo Bay prison	34.84	3.42	2.38	2.33	502
15. UN elects female Secretary General	47.37	3.01	2.02	2.04	472
16. Terrorist attack kills >100 Americans	38.12	3.23	2.30	2.57	511
17. Russia's economy contracts in 2016	58.14	3.13	2.19	1.91	465
18. Ayman al-Zawahiri captured or killed	37.44	2.73	2.07	2.06	511
19. Snowden returns to United States	26.34	3.20	2.37	2.33	494
20. More than 10 U.S. soldiers killed in Iraq in 2016	68.26	3.49	2.47	2.34	511
Mean	*48.03*	*3.40*	*2.29*	*2.34*	*10,000*
Standard deviation	*27.29*	*1.19*	*1.10*	*1.22*	*10,000*

Table S5 presents mean assessments of uncertainty for each question that our survey posed. Full question wordings are provided earlier in this supplement. Confidence levels were elicited on 5-point scales. Shading in Table S5 reflects the nearest integer to which mean confidence assessments would be rounded.

Table S5. Assessments of uncertainty across questions

8. Full analysis of results from survey experiment 2 (ordinary least squares)

Table S6 presents full results for the analysts of Survey Experiment 2.

	Model 1: Reliability of Available Evidence	*Model 2:* Range of Reasonable Opinion	*Model 3:* Responsiveness to New Information
1. Hillary Clinton elected president	0.36 (.05)***	-0.47 (.06)***	0.63 (.06)***
2. Garland confirmed to Supreme Court†	-0.09 (.06)	-0.09 (.05)	0.17 (.05)**
3. Obama approval rating over 50%†	0.20 (.05)***	-0.42 (.06)***	0.37 (.06)***
4. Unemployment rate under 5%†	0.16 (.05)**	-0.14 (.05)**	0.01 (.05)
5. Washington Redskins trademark restored†	-0.28 (.06)***	-0.16 (.05)**	0.21 (.06)***
6. Third-party wins 10% presidential vote†	0.08 (.06)	-0.16 (.06)**	0.14 (.06)*
7. Congress passes any gun control measure†	0.27 (.05)***	-0.20 (.06)***	0.29 (.06)***
8. Avg. gasoline price under $2.50/gal†	0.18 (.05)***	-0.13 (.05)*	-0.09 (.05)
9. Bergdahl sentenced to ≥5 years prison	-0.25 (.06)***	-0.06 (.05)	-0.07 (.05)
10. Avg. temp. higher in 2016 than 2015	0.44 (.05)***	0.10 (.06)	0.16 (.06)**
11. Britain formally exits European Union†	0.24 (.05)***	-0.14 (.06)*	0.10 (.06)
12. Bashar al-Assad no longer in power†	-0.32 (.06)***	-0.11 (.05)*	-0.03 (.05)
13. U.S. admits ≥7,500 Syrian refugees†	0.00 (.06)	-0.17 (.05)**	0.03 (.05)
14. Guantanamo Bay prison camp closed†	-0.07 (.05)	-0.08 (.05)	0.03 (.05)
15. UN elects a female Secy. General	-0.15 (.06)**	-0.20 (.05)***	0.11 (.05)**
16. A single terrorist event kills ≥100 in U.S.†	-0.30 (.06)***	-0.23 (.05)***	0.35 (.05)***
17. Russia's economy contracts in 2016	-0.08 (.06)	0.01 (.05)	-0.12 (.05)*
18. Ayman al-Zawahiri captured or killed†	-0.58 (.06)***	-0.19 (.05)***	0.09 (.05)
19. Edward Snowden returns to United States†	-0.34 (.06)***	-0.09 (.05)	0.03 (.05)
20. ≥10 U.S. soldiers killed in Iraq in 2016	-	-	-
Certainty (0.0-0.5)	2.09 (.07)***	0.65 (.07)***	1.15 (.08)***
Reliability of available evidence (1-5)	-	0.08 (.01)***	0.16 (.01)***
Range of reasonable opinion (1-5)	0.10 (.01)***	-	0.45 (.01)***
Responsiveness to new information (1-5)	0.17 (.01)***	0.41 (.01)***	-
Female (0,1)	-0.08 (.04)*	-0.01 (.03)	0.06 (.03)
White (0,1)	0.04 (.04)	-0.14 (.04)***	-0.06 (.04)
College-educated (0,1)	0.01 (.04)	-0.02 (.03)	-0.02 (.03)
Liberalism (1-7)	-0.02 (.01)	0.00 (.01)	-0.05 (.01)***
Age (integer)	$5.8e^{-4}$ ($1.7e^{-3}$)	$3.6e^{-3}$ ($1.3e^{-3}$)**	$3.6e^{-3}$ ($1.5e^{-3}$)***
Political engagement (integer)	0.02 ($2.7e^{-3}$)***	$4.9e^{-3}$ ($2.2e^{-3}$)*	$5.3e^{-3}$ ($3.0e^{-3}$)
Constant	2.27 (.11)***	1.31 (.09)***	0.45 (.10)
R^2	0.27	0.31	0.37
N	9,870	9,870	9,870

Table S6 presents ordinary least squares regressions modeling variation in how respondents assigned confidence levels across survey questions. All models include respondent fixed effects and standard errors clustered by respondent. See supplementary material for full question wordings, alternative specifications, and robustness checks. *: p<0.05, **: p<0.01, ***: p<0.001. †: "… by [or at] the end of 2016."

Table S6. Exploring independent variation across confidence levels

9. Full analysis of results from survey experiment 2 (ordered logit)

Table S7 replicates our analysis of results from survey experiment 2 using ordered logit. The only substantive changes from what we present in the main text of the paper are that (i) the coefficient for Question 12 in Model 2 loses statistical significance (its *p*-value is now 0.10), and (ii) the coefficient for Question 18 in Model 3 becomes statistically significant.

	Model 1: Reliability of Available Evidence	*Model 2:* Range of Reasonable Opinion	*Model 3:* Responsiveness to New Information
1. Hillary Clinton elected president	0.84 (.13)***	-1.16 (.06)***	1.42 (.13)***
2. Garland confirmed to Supreme Court	-0.16 (.13)	-0.18 (.12)	0.45 (.13)**
3. Obama approval rating over 50%	0.42 (.12)***	-1.00 (.14)***	0.91 (.13)***
4. Unemployment rate under 5%	0.37 (.12)**	-0.32 (.13)**	0.04 (.13)
5. Washington Redskins trademark restored	-0.60 (.13)***	-0.34 (.12)**	0.57 (.13)***
6. Third-party wins 10% presidential vote	0.17 (.13)	-0.39 (.12)**	0.38 (.13)**
7. Congress passes any gun control measure	0.69 (.12)***	-0.48 (.13)***	0.67 (.12)***
8. Avg. gasoline price under $2.50/gal	0.41 (.12)***	-0.29 (.13)*	-0.14 (.13)
9. Bowe Bergdahl sentenced to ≥5 years prison	-0.55 (.12)***	-0.10 (.12)	-0.13 (.13)
10. Avg. global temp. higher in 2016 than 2015	1.17 (.13)***	0.23 (.13)	0.36 (.13)**
11. Britain formally exits European Union	0.64 (.13)***	-0.33 (.13)*	0.18 (.13)
12. Bashar al-Assad no longer in power	-0.64 (.12)***	-0.20 (.12)	-0.04 (.12)
13. U.S. admits ≥7,500 Syrian refugees	-0.02 (.13)	-0.38 (.13)**	0.12 (.13)
14. Guantanamo Bay prison camp closed	-0.19 (.12)	-0.17 (.13)	0.15 (.12)
15. UN elects a female Secy. General	-0.31 (.12)*	-0.47 (.13)***	0.33 (.13)*
16. A single terrorist event kills ≥100 in U.S.	-0.61 (.13)***	-0.52 (.13)***	0.86 (.13)***
17. Russia's economy contracts in 2016	-0.15 (.13)	0.05 (.12)	-0.36 (.13)**
18. Ayman al-Zawahiri captured or killed	-1.24 (.13)***	-0.42 (.12)***	0.27 (.13)*
19. Edward Snowden returns to United States	-0.73 (.13)***	-0.20 (.13)	0.12 (.12)
20. ≥10 U.S. soldiers killed in Iraq in 2016	-	-	-
Certainty (0.0-0.5)	5.33 (.18)***	1.35 (.18)***	2.32 (.19)***
Reliability of available evidence (1-5)	-	0.22 (.03)***	0.43 (.03)***
Range of reasonable opinion (1-5)	0.30 (.03)***	-	1.07 (.04)***
Responsiveness to new information (1-5)	0.47 (.03)***	1.00 (.04)***	-
Female (0,1)	-0.23 (.10)*	-0.02 (.08)	0.14 (.09)
White (0,1)	0.13 (.12)	-0.35 (.10)***	-0.19 (.11)
College-educated (0,1)	0.04 (.10)	-0.07 (.08)	-0.05 (.09)
Liberalism (1-7)	-0.03 (.03)	0.01 (.02)	-0.14 (.03)***
Age (integer)	1.4e^{-3} (4.2e^{-3})	-0.01 (3.6e^{-3})*	8.2e^{-3} (3.9e^{-3})*
Political engagement (integer)	0.04 (.01)***	1.2e^{-2} (6.1e^{-3})*	0.01 (.01)
Cut points	-0.48, 1.36, 2.75, 5.73	0.51, 3.28, 4.67, 6.85	2.86, 5.34, 6.59, 8.71
N	9,870	9,870	9,870

Table S7 presents ordered logit models predicting variation in how respondents assigned confidence levels across survey questions. All models include respondent fixed effects and standard errors clustered by respondent. *: p<0.05, **: p<0.01, ***: p<0.001. Full question wordings are provided earlier in this supplement.

Table S7. Exploring independent variation across confidence levels (ordered logit)

10. Relationships between respondent demographics and levels of confidence

Table S8 explores correlations between respondent demographics and levels of confidence, measured on 7-point scales.[1] We ordered each scale so that higher values indicate great confidence (e.g., more reliable evidence, less range of reasonable opinion, less responsiveness to new information). We control for the *Certainty* that respondents assigned to their probability estimates (i.e. the absolute value of the difference between their probability estimates and 0.50) so as to isolate variations in confidence levels as distinct from how respondents assessed probability. We also include a model showing relationships between demographic variables and *Certainty* itself. All models are ordinary least squares with standard errors clustered by respondent. Since respondents for this survey experiment were recruited via Amazon Mechanical Turk, the results of this analysis are purely exploratory.

The most consistent finding across these models is that respondents who were more liberal (with liberalism measured on the standard 7-point scale used by the American National Election Survey) tended to place less confidence in their assessments of uncertainty. This finding is statistically significant at the $p<0.05$ level for all three conceptions of confidence. We also found that respondents who reported consuming more political news believed that they possessed more reliable evidence and that their views were less likely to change in response to new information.

[1] Note that these models differ from those presented in Table S7 because they do not control for other kinds of confidence that analysts assessed. Thus if one category of respondents systematically assigned greater levels to all three kinds of confidence at once, this would *not* lead to statistically significant coefficients in Table S7, but it *would* lead to statistically significant coefficients in Table S8.

White respondents tended to place less confidence in their judgments than other respondents, but they also attached greater levels of certainty to their probability estimates. Gender, age, and college education bore no consistent relationships to any of the three kinds of confidence that we asked respondents to assess.

	Model 1: Reliability of available evidence	*Model 2:* Range of reasonable opinion	*Model 3:* Responsiveness to new information	*Model 4:* Certainty
Female	-0.08 (.04)	0.00 (.04)	0.05 (.04)	$-3.6e^{-3}$ ($4.7e^{-3}$)
White	-0.02 (.05)	-0.20 (.05)***	-0.16 (.05)**	0.02 (.01)**
Age	$8.5e^{-4}$ ($1.8e^{-3}$)	$-2.4e^{-3}$ ($1.6e^{-3}$)	$-2.6e^{-3}$ ($1.9e^{-3}$)	$3.7e^{-4}$ ($1.9e^{-4}$)*
College education	0.01 (.04)	-0.03 (.04)	-0.03 (.04)	$6.7e^{-3}$ ($4.6e^{-3}$)
Hours per week of news consumed	0.02 (.00)***	$2.2e^{-4}$ ($2.7e^{-3}$)	$8.3e^{-3}$ ($3.5e^{-3}$)*	$1.3e^{-3}$ ($3.4e^{-4}$)***
Liberalism	-0.03 (.01)**	-0.02 (.01)*	-0.06 (.01)***	$6.1e^{-4}$ ($1.3e^{-3}$)
Certainty	2.93 (.09)**	1.91 (.10)***	2.60 (.10)***	-
Constant	2.77 (.10)***	2.24 (.10)***	2.01 (.11)***	0.19 (.01)***
N	9,870	9,870	9,870	9,870
R^2	0.17	0.08	0.12	0.15

Table S8 presents ordinary least squares regressions exploring relationships between respondent demographics and assessments of uncertainty. Standard errors clustered by respondent. *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

Table S8. Respondent demographics and analytic confidence

Political Psychology

Political Psychology, Vol. 39, No. 5, 2018
doi: 10.1111/pops.12465

Analytic Confidence and Political Decision-Making: Theoretical Principles and Experimental Evidence From National Security Professionals

Jeffrey A. Friedman
Dartmouth College, Institute for Advanced Study in Toulouse

Richard Zeckhauser
Harvard University

When making decisions under uncertainty, it is important to distinguish between the probability that a judgment is true and the confidence analysts possess in drawing their conclusions. Yet analysts and decision-makers often struggle to define "confidence" in this context, and many ways that scholars use this term do not necessarily facilitate decision-making under uncertainty. To help resolve this confusion, we argue for disaggregating analytic confidence along three dimensions: reliability of available evidence, range of reasonable opinion, and responsiveness to new information. After explaining how these attributes hold different implications for decision-making in principle, we present survey experiments examining how analysts and decision-makers employ these ideas in practice. Our first experiment found that each conception of confidence distinctively influenced national security professionals' evaluations of high-stakes decisions. Our second experiment showed that inexperienced assessors of uncertainty could consistently discriminate among our conceptions of confidence when making political forecasts. We focus on national security, where debates about defining "confidence levels" have clear practical implications. But our theoretical framework generalizes to nearly any area of political decision-making, and our empirical results provide encouraging evidence that analysts and decision-makers can grasp these abstract elements of uncertainty.

KEY WORDS: decision-making, uncertainty, national security, intelligence, experiments

This article explores the distinction between probability and confidence in political decision-making. Broadly speaking, probability reflects an analyst's estimate of the chances that a statement is true, while confidence reflects the degree to which an analyst believes that he or she possesses a sound basis for assessing uncertainty. These concepts can vary independently, and each represents a distinct element of high-stakes decision-making. In U.S. criminal courts, for example, jurors should not send defendants to jail just because their guilt seems to be likely. In principle, guilty verdicts require concrete evidence, not just subjective certainty. The criminal justice system thus requires jurors to assess both high probability and high confidence to find for conviction.

The relationship between probability and confidence shapes many political issues. Scholars often argue that the use of force in international relations, like the removal of civil liberties in a court of

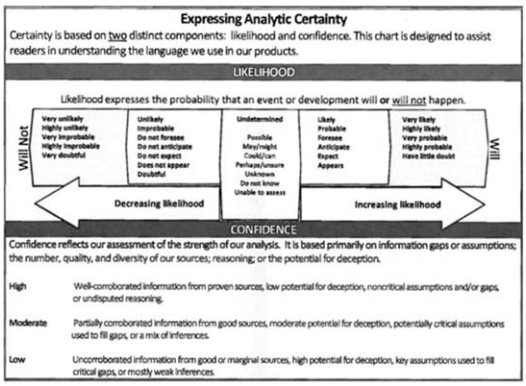
law, must be justified through reliable evidence and not merely the belief that coercion will produce favorable results. When scholars and pundits debate election forecasts, they are (at least implicitly) arguing *both* about what probabilities they should assign to electoral outcomes *and* about how much confidence those estimates deserve. Though virtually all scientists believe that carbon emissions are harming the global climate, skeptics observe that scientists lack a commonly accepted framework for modeling climate change, and therefore they argue that these predictions are insufficiently reliable to justify costly mitigation efforts. In this way, political judgment requires assigning probabilities to significant events as well as judging the quality of those estimates before their accuracy can be known. This is the sense in which we use the term "confidence" (or "analytic confidence") in this article.

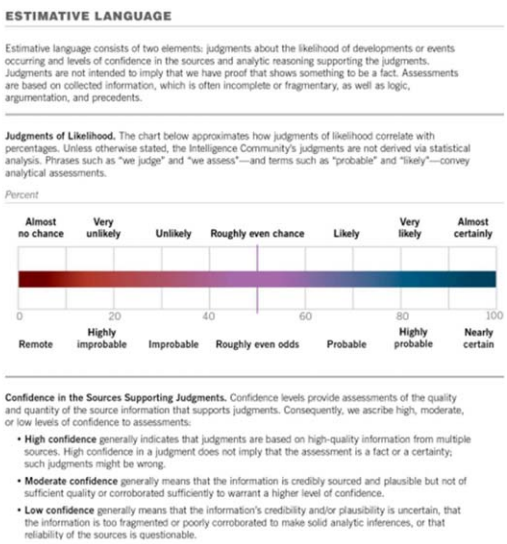
Political scientists have said relatively little about analytic confidence in this context. Indeed, most of the ways in which the term "confidence" appears in contemporary scholarship have little to do with judging the extent to which analysts believe they possess a sound basis for making a particular assessment of uncertainty. Some scholars define confidence as an attribute belonging to individuals, such as how precisely analysts tend to estimate uncertain quantities or how highly they rate their overall performance (Moore & Healy, 2008; Ortoleva & Snowberg, 2015a). Others define confidence as an attribute of available data, such as how "confidence intervals" characterize statistical parameters. Still others define confidence based on ex post evaluations of judgmental accuracy, such that an analyst is "overconfident" if she assigns excessive certainty to her judgments (Johnson, 2004; Tetlock, 2005). When survey researchers ask respondents to say how "confident" they are in their political knowledge, this is often equivalent to eliciting beliefs about the chances that those statements are true.

The fact that the word "confidence" conveys so many distinct ideas provides a pragmatic reason to employ clearer terminology when debating high-stakes issues. But this article also provides a theoretical argument for why traditional discussions of analytic confidence are excessively vague. In particular, we explain how there are at least three distinct ways of describing the extent to which an analyst believes that he or she possesses a sound basis for assessing uncertainty: reliability of available evidence, range of reasonable opinion, and responsiveness to new information. We distinguish these attributes from probability estimates, explain how they can vary independently, and argue that each carries different implications for political decision-making. We focus this discussion on national security, as this is a domain where scholars have long debated conceptual frameworks for opening up the "black box" of uncertainty (Rathbun, 2007; Vertzberger, 1995), and where these debates directly shape the production of intelligence reports, military plans, and other consequential analyses (Fingar, 2011; Nye, 1994). Yet we emphasize throughout the article how our three conceptions of confidence apply to nearly any other area of political decision-making where uncertainty plays a role.

After developing a theoretical framework for understanding the relationship between analytic confidence and political decision-making in principle, we present two survey experiments examining the extent to which respondents can employ these ideas in practice. Our first survey experiment shows how a cross-section of national security elites, recruited through the National War College, employed all three conceptions of confidence simultaneously when evaluating decisions under uncertainty. Our second experiment demonstrates that even novice analysts, recruited from Amazon Mechanical Turk, could consistently discriminate among our three conceptions of confidence when making political forecasts.

Beyond supporting the validity of our conceptual framework, these findings speak to broader academic questions about the ability of political analysts and decision-makers to employ abstract reasoning when assessing uncertainty. Behavioral studies often find that respondents' perceptions of uncertainty are more consistent with affect-laden feelings than with reasoned judgments (Gigerenzer, 2008; Kahneman, 2011, 201; Loewenstein, Hsee, Weber, & Welch, 2001; Slovic, 2010). This skepticism is particularly strong in national security, where many scholars view the challenge of assessing uncertainty as being unusually complex and subjective (Betts, 2000; Jervis, 1997). We nevertheless demonstrate that both elite and nonelite respondents can consistently distinguish between probability





2015 Defense Intelligence Agency Guidelines *2017 National Intelligence Council Guidelines*

Figure 1. Instructions for expressing uncertainty in intelligence. [Color figure can be viewed at wileyonlinelibrary.com]

and confidence and that they can reliably divide assessments of confidence into three constituent parts. In this respect, our findings provide encouraging evidence regarding the capacity of analysts and decision-makers to engage with abstract elements of uncertainty when given the opportunity to do so.

Three Conceptions of Confidence

The challenge of disentangling probability and confidence is prominent in the domain of intelligence and national security. Figure 1 presents guidelines on this subject from the U.S. Defense Intelligence Agency and National Intelligence Council. Though these standards instruct analysts to convey probability estimates using qualitative language, it is at least clear that these phrases correspond to numbers between 0% and 100%. By contrast, these guidelines define confidence with respect to attributes such as quality and quantity of available information, potential for deception, gaps in knowledge, and strength of relevant inferences. It is hard to say how analysts should operationalize any of these attributes, let alone how one might represent those factors on a single scale. Vagueness surrounding confidence levels thus extends beyond ambiguous language: Interpreting these assessments requires grappling with conceptual confusion about what analytic confidence means.

This confusion has two main elements. First, it is unclear why some factors associated with analytic confidence should matter for decision-making. For example, both the Defense Intelligence Agency and the National Intelligence Council define confidence with respect to the number of sources that analysts use to draw conclusions. But if an analyst finds a single, dispositive source, then there is no reason why she cannot make a sound judgment, just as one credible eyewitness can justify a criminal conviction. Thus, one challenge in defining analytic confidence is understanding which factors represent primary elements of uncertainty and which proxy for other, more important concepts. In this respect, our theoretical framework hones discussions of confidence into three basic ways of describing the extent to which analysts believe they possess a sound basis for assessing uncertainty. We call these attributes reliability of available evidence, range of reasonable opinion, and responsiveness to new information.

Existing discussions of analytic confidence also tend to be underspecified. For example, when the Defense Intelligence Agency describes "strength of analysis" or when the National Intelligence Council discusses "high-quality information" (see Figure 1), these phrases suggest constellations of factors as opposed to precise ideas. In describing our three conceptions of confidence, we thus explain how each can be quantified in principle. This is not to claim that analysts *should* quantify these attributes in practice, as debates about quantifying subjective judgments are an important topic of controversy in their own right (Kent, 1964). Instead, we specify each conception of confidence to demonstrate that these ideas reflect coherent and distinct elements of uncertainty.

Saying that these concepts are "distinct" does not require them to be uncorrelated. Probability and confidence often go together, as analysts who possess greater certainty also tend to possess greater confidence. Throughout this section, however, we explain how this correlation is imperfect, and we offer a range of examples to demonstrate how separating these elements of uncertainty can play an important role in shaping high-stakes decisions. In conducting this discussion, we frequently rely on the concept of "second-order uncertainty," which captures the relative credibility that analysts assign to every possible judgment. Second-order uncertainty thus reflects an analyst's belief about the chances that the "right" probability estimate to make on the basis of available evidence is 0%, 1%, 2%, and so on up to 100%. Each of our three conceptions of confidence essentially articulates a different way in which second-order uncertainty shapes political decision-making.

Of course, we cannot say that reliability of available evidence, range of reasonable opinion, and responsiveness to new information are the *only* ways of describing analytic confidence that would *ever* matter to making political decisions. Yet our three conceptions of confidence cover each basic element of Bayesian reasoning. Reliability of available evidence reflects the degree to which an analyst's judgments depend on her initial prior assumptions, range of reasonable opinion captures an analyst's current perception of second-order uncertainty, and responsiveness to new information reflects how further study might shift those perceptions into what decision theorists call "posterior" judgments. At the very least, we will argue that our three conceptions of confidence *should* matter for making *all* political decisions where uncertainty plays a role. To our knowledge, our theoretical framework reflects the first attempt to specify the foundations of analytic confidence in this manner.

Reliability of Available Evidence

Our first conception of analytic confidence is reliability of available evidence. In this view, analysts have a sound basis for assessing uncertainty when their judgments reflect case-specific knowledge as opposed to speculation. This is the most common way that scholars and practitioners define confidence levels in intelligence studies. Thus, Thomas Fingar (2011), formerly Deputy Director of National Intelligence for Analysis, writes that "confidence judgments are based on the quantity, quality, and consistency of the information available" (p. 88; Wheaton, 2012, p. 335). Assessing the reliability of available evidence requires considering factors such as the volume of available information, the diagnostic value of individual data points, analysts' subject matter expertise, and the extent to which independent sources provide corroborating views. We can summarize these factors with respect to a single attribute, which is the extent to which an analyst's judgment depends on her initial prior beliefs.[1]

While judgments based on more reliable evidence often entail greater certainty, those attributes can vary independently. For example, estimates of the chances that a card drawn from a randomly shuffled deck will be black (one-in-two), a diamond (one-in-four), or an ace (one-in-thirteen) all reflect equally reliable reasoning. Identical assessments of confidence can thus accompany varying

[1] We can quantify this attribute as the statistical difference between an analyst's second-order uncertainty and what that judgment would have been if her initial prior assumptions were uninformed (i.e., if she initially assigned equal credibility to every relevant hypothesis). This captures the degree to which the analyst's initial prior assumptions continue to shape her assessment of uncertainty after viewing case-specific evidence.

estimates of probability. Identical probability estimates can also accompany varying degrees of confidence. Thus, if a political analyst says there is a 50% chance that a candidate will win an election, this could reflect the result of a highly rigorous study showing that the race is even. But this analyst's lack of certainty could also reflect that she has *not* studied the election in detail and therefore possesses no reason to believe that one candidate is more likely to win than the other.

It is furthermore possible for analysts to possess high certainty without the benefit of reliable evidence. This is why courts of law instruct juries to avoid reaching verdicts based on stereotypes, hearsay, or particular kinds of circumstantial evidence. The motivation behind these restrictions is the concern that jurors can become convinced of a defendant's guilt even in the absence of reliable evidence to support that judgment (McAuliff, 1982). A similar premise guides the jury selection process, on the assumption that prejudicial jurors might reach conclusions that are heavily influenced by prior beliefs as opposed to rigorous, case-specific reasoning. A different example of assessing high certainty without reliable evidence would be a doctor who tells a patient it is very unlikely that she possesses a rare disease simply because the base rate of occurrence for that disease is small. Such reasoning is distinct from making a low-probability diagnosis after running a large battery of tests that search for evidence of a disease in this particular patient. In the latter case, the doctor would then possess a high degree of certainty *and* the ability to defend that judgment without relying on her prior assumptions.

In classical decision theory, the reliability of available evidence does not directly shape decision-making. Strictly from the standpoint of expected utility, it does not matter whether a probability estimate is *known* to take a certain value or whether it is *expected* to take that value on the basis of ambiguous evidence (Ellsberg, 1961). Though many individuals would not treat these cases equally, most decision theorists see such "ambiguity aversion" as misguided (Al-Najjar & Weinstein, 2009).

Reliability of available evidence nevertheless carries important implications for the ethics of political decision-making. In particular, many people believe that it is inappropriate for public policy to cause harm without concrete justification. Controversy over racial profiling by police is, thus, largely a debate over whether it is acceptable to impose costs on individual citizens without reliable indicators of personal guilt. Similarly, when debating environmental regulations, drug approvals, and proposals to privatize social security, decision-makers must consider the extent to which they can prove that their actions do not impose unnecessary risks.[2] In this sense, reliability of available evidence shapes political choices in a manner that probability estimates alone cannot capture.

George W. Bush's presidency offers two demonstrations of how similar concerns surround national security decision-making. The first case is the 2003 invasion of Iraq, which shows the danger of failing to draw clear distinctions between probability and confidence. The Bush administration publicly justified the war as a measure to preempt Saddam Hussein from obtaining nuclear weapons. That claim relied on circumstantial evidence interpreted in light of Saddam's past behavior. Many scholars believe that the speculative nature of this analysis undermined the Bush administration's appeal to preemption, as invoking that doctrine requires reliable evidence of imminent threat (Doyle, 2008). In this view, one of the main flaws with prewar intelligence on Iraq was not just that analysts mistakenly concluded Saddam was likely pursuing nuclear weapons, but that published reports failed to clarify the degree to which this conclusion relied on prior beliefs (Betts, 2007, p. 116; Jervis, 2006, p. 44). According to former Deputy Director of Intelligence Michael Morell, "By far the biggest mistake made by the analysts. . . was not that they came to the wrong conclusion about Iraq's WMD program, but rather that they did not rigorously ask themselves how confident they were in their judgments" (Morell, 2015, p. 102).

A second demonstration of the importance of drawing such distinctions occurred later in Bush's presidency, in response to reports that Syria was building a nuclear reactor at a site called al-Kibar. Former CIA Director Michael Hayden (2016) describes briefing President Bush about this subject in

[2] Such reasoning is especially prominent in debates over the precautionary principle.

2007, explaining that while analysts were virtually certain that al-Kibar was a military facility, they possessed no reliable evidence supporting this claim. "This is part of a Syrian weapons program. Of course it is," Hayden recalls saying. Yet he continued, "I can't find the other parts of the weapons program. No reprocessing facility. No weaponization effort that we can see. So I can only give this to you with low confidence." Hayden then recounts how "the president observed that his preemption policy demanded a threat be imminent before we could act. Our estimate of low confidence in a weapons program made that very difficult to justify, and therefore, the president declared, 'We will *not* strike the facility'" (p. 262). This statement highlights how subjective probability estimates alone do not capture all elements of uncertainty that inform high-stakes decisions. Reliability of available evidence also plays an important role in shaping the ethical and political context for taking action.

Range of Reasonable Opinion

Our second conception of confidence is range of reasonable opinion. In this view, analysts have a sounder basis for assessing uncertainty when they believe there is a narrower set of plausible viewpoints. Thus, range of reasonable opinion resembles the way that social scientists describe "confidence intervals," though given how most statistical models rely on debatable assumptions, it is usually a mistake to believe that statistical output objectively characterizes the ambiguity surrounding a given inference (Manski, 2013). Analysts can quantify the way they perceive the range of reasonable opinion by describing their second-order uncertainty as to what the right probability estimate entails.

In statistics, larger volumes of data typically produce tighter confidence intervals. By the same logic, it is intuitive to see range of reasonable opinion as closely connected to the reliability of available evidence. But it is a mistake to conceive of those attributes as identical. The key to this distinction is that reliability of available evidence describes how analysts reach their conclusions (in particular, the extent to which those conclusions depend on analysts' initial prior assumptions). Range of reasonable opinion, by contrast, describes the amount of ambiguity analysts are willing to accept about what the right conclusion entails. Thus, if an analyst holds an extremely strong prior, then she could believe the correct judgment is unambiguous, even if she cannot justify that conclusion on the basis of case-specific evidence. Studies suggest that this is indeed how most people process political information, relying on cues provided by experts whom they generally trust rather than forming rigorous, case-specific judgments (Achen & Bartels, 2016; Zaller, 1992).

On the flip side, even when analysts base their conclusions entirely upon rigorous, case-specific reasoning, they can still see the evidence as being objectively ambiguous. For instance, when climate scientists estimate the chances that global temperatures will rise by a given amount or the chances that sea levels will rise to a given level, these judgments generally reflect the output of detailed, rigorous simulations. Yet scientists who rely on different data sets or different models often reach substantially different conclusions. In this context, we would say that climate scientists' assessments of uncertainty reflect high degrees of reliable evidence (at least in the sense that these judgments leave little room for analysts' prior assumptions), but that they still permit a substantial range of reasonable opinion. As shown in Figure 2, the Intergovernmental Panel on Climate Change (IPCC) explicitly recognizes this distinction, asking analysts to describe the quality of available information *and* the extent of expert agreement when assessing analytic confidence (Mastrandrea et al., 2010, p. 3).

Range of reasonable opinion shapes political decisions in three main ways. First, as with reliability of available evidence, low levels of confidence in this area raise ethical concerns. In most U.S. criminal trials, juries can only convict defendants by unanimous vote. A criminal trial's outcome thus depends on assessing probability (guilt beyond "reasonable doubt"), reliability of evidence (such as the exclusion of hearsay), and range of reasonable opinion (via jury consensus). The broader practice of establishing veto power in political decision-making serves a similar purpose, and even when unanimity is not formally required to approve a policy, many organizations strive to achieve consensus to maximize the legitimacy of their actions.

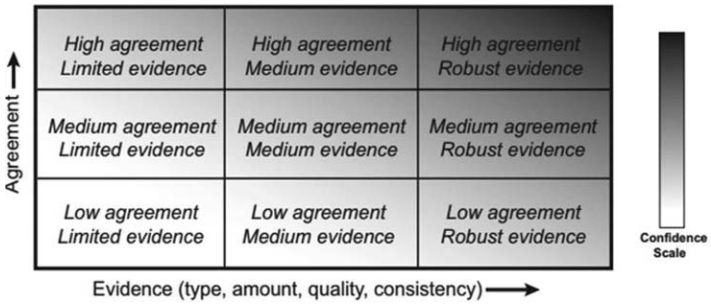


Figure 2. IPCC guidelines for assessing confidence.

The Iraq War debate shows how range of reasonable opinion can also shape the politics of high-stakes decision-making. The 2002 National Intelligence Estimate on Iraq's weapons of mass destruction (WMD) programs contained a dissent from the U.S. Department of State. State Department analysts disputed the claim that Iraq's acquisition of aluminum tubes indicated its intent to build uranium centrifuges. Though scholars disagree over how much weight this dissent deserved, many observers argue that it was inappropriate for the U.S. Intelligence Community to offer such a high-stakes judgment, or for the Bush administration to act on that judgment, without conclusively rebutting the State Department's dissent (U.S. Senate, 2005, pp. 87–118).

Describing the range of reasonable opinion can also be valuable in prompting decision-makers to explore the basis for existing disagreement. During the search for Osama bin Laden, for example, President Obama asked his advisers to quantify the chances that al-Qaeda's leader was living in Abbottabad, Pakistan. The head of the Central Intelligence Agency's bin Laden team placed these chances at 95% while CIA Deputy Director Michael Morell put his own guess at 60%. When President Obama asked about this discrepancy, Morell explained that while counterterrorism officials trusted their targeting abilities given recent tactical successes, his experience assessing Iraq's weapons of mass destruction programs had left him wary of basing strategic decisions on circumstantial evidence. Morell (2015, pp. 156–161) recounts how this discussion played an important role framing President Obama's choice over whether to strike the Abbottabad compound. For similar reasons, Richard Neustadt and Ernest May (1986) recommend that decision-makers ask analysts to quantify probability estimates when debating major decisions, as "We know of no better way to force clarification of meanings while exposing hidden differences" (p. 152).

Responsiveness to New Information

Our third conception of confidence is responsiveness to new information. In this view, individuals should be more "confident" when assessing uncertainty the less they expect subsequent analysis to change their judgments. Responsiveness to new information depends both on how strongly analysts hold their beliefs and how much information they might be able to collect if given additional time and resources (Friedman & Zeckhauser, 2015). Formally, we can express these factors' combined influence in terms of second-order uncertainty about the value a probability estimate might take at some future date.

Responsiveness to new information differs from our other conceptions of confidence in that it cannot be divorced from practical considerations. Analysts can assess reliability of available evidence and range of reasonable opinion without referencing the decisions they are trying to inform, as these attributes rely solely on the way analysts perceive uncertainty and how they arrived at their judgments. By contrast, there is no way to assess how these judgments might change moving forward without making assumptions about the resources that decision-makers can devote to further analysis. The more decision-makers are willing to invest, the more likely such investments will be to shift analysts'

judgments. In other words, the key issue when assessing responsiveness to new information is not how much analysts might be able to reduce remaining uncertainty in principle, but rather the extent to which they expect their views to change in practice given available constraints.

Neither reliability of available evidence nor range of reasonable opinion addresses this subject directly. For example, some of the most important questions in political discourse are also highly subjective. What is the risk of another global financial crisis within the next decade? What are the chances that a new Supreme Court justice would vote to overturn *Roe v. Wade*? What is the probability that the United States and China will go to war by 2030? Any answer to these questions will rely heavily on speculation while leaving wide room for reasonable disagreement. We might, thus, consider those judgments to involve "low confidence." Yet seen through the lens of responsiveness to new information, these questions often provide the *strongest* basis for assessing uncertainty, given that further study is unlikely to change analysts' views within a relevant time frame.

Drawing such distinctions is especially important when determining how to time major decisions. For example, when intelligence officials first approached President Obama with the idea that Osama bin Laden might be living in Abbottabad in fall 2010, Obama asked them to pursue alternative methods for collecting information about the compound. Though some of these methods yielded valuable information, none conclusively identified the compound's occupants (Morell, 2015, p. 152). By spring 2011, it was clear that intelligence officials had exhausted feasible avenues for collecting actionable evidence. Meanwhile, the longer President Obama waited to authorize a raid on the compound, the greater the chances became that the compound's occupants would learn they were being watched or that they would leave for other reasons. Regardless of the reliability of available evidence about the Abbottabad compound or the range of reasonable opinion surrounding this subject, President Obama therefore determined that he could not justify further delay. Responsiveness to new information was thus the only major element of uncertainty that shifted between fall 2010 and spring 2011—but that shift also proved decisive in shaping President Obama's decision to strike Abbottabad.

President Bush's decision not to attack Syria's nuclear reactor in 2007 also highlights the importance of weighing responsiveness to new information. Though all analysts agreed that the al-Kibar facility was likely designed for military purposes, they believed that the reactor was years away from producing weapons-grade material. There was thus little opportunity cost in continuing to study the facility. Meanwhile, intelligence officials had viable opportunities to continue gathering information about the site, and it was possible that the situation would be resolved without U.S. intervention. (Indeed, Israel destroyed the facility with an air strike later that year.) Compared to the bin Laden raid, analysts had higher certainty, more reliable evidence, and greater analytic consensus in assessing al-Kibar. Yet President Bush chose not to strike the Syrian reactor, in part because gathering additional information was more attractive than striking immediately.

Testing Conceptual Validity Through Survey Experiments

The previous section argued that the common practice of ascribing "confidence levels" to assessments of uncertainty obscures information that is relevant to making high-stakes decisions. We argued that analytic confidence is not a single variable, but rather three distinct attributes that can vary independently, each holding different implications for political decision-making.

Yet explaining how our three conceptions of confidence operate in principle does not guarantee that these distinctions matter in practice. Behavioral researchers often find that analysts and decision-makers struggle to manage abstract elements of uncertainty. In many areas of high-stakes decision-making, perceptions of uncertainty seem more consistent with affect-laden feelings than with reasoned judgments (Gigerenzer, 2008; Kahneman, 2011, p. 201; Loewenstein et al., 2001; Slovic, 2010). This presents two main challenges to our theoretical framework. First, decision-makers may not reliably respond to distinctions among our three conceptions of confidence, in which case our ideas would add

little *descriptive* value for understanding how political decisions are made. Second, analysts might be unable to discriminate among our three conceptions of confidence, implying that our ideas hold little *prescriptive* value for improving assessments of uncertainty in national security or other fields.

We conducted two survey experiments to address these concerns. The first shows how a cross-section of national security elites, recruited through the National War College, employed all three conceptions of confidence simultaneously when evaluating decisions under uncertainty. Our second survey experiment demonstrates that novice political analysts, recruited through Amazon's Mechanical Turk survey platform, could consistently discriminate among our three conceptions of confidence when making political forecasts. These findings suggest that our attempts to open up the "black box" of analytic confidence yield both descriptive and prescriptive benefits: We see how these distinctions shape the way that real national security officials evaluate high-stakes decisions, and we find that analysts can intuitively grasp these distinctions.

Survey experiments have four main drawbacks for the study of political decision-making. First, survey experiments cannot replicate the dense mixture of cues and pressures that real high-stakes decisions entail. This raises well-known concerns about survey experiments' external validity (Hyde, 2015). It is particularly problematic to use survey experiments to draw inferences about the overall levels of support which people might offer to real political decisions that involve many other factors. The strength of the experimental method instead lies in examining how varying specific elements of a decision-making problem systematically influences respondents' reactions. This ability to isolate specific microfoundations of analysis and decision-making is especially important for the purposes of this article given how national security officials almost always leave their assessments of uncertainty vague, and given that they rarely disentangle different conceptions of confidence in the way this article proposes.[3] Indeed, the main purpose of these experiments is to determine whether it *would* be valuable for national security officials to parse these judgments in a more detailed manner. In these respects, survey experiments provide leverage that observational studies cannot provide when it comes to understanding how specific conceptions of confidence shape analysis and decision-making.

A related drawback with survey experiments is that it is difficult to calibrate the strength of experimental manipulations (Barabas & Jerit, 2010). Simulating what informational "treatments" might look like in real decision-making scenarios inevitably sacrifices realism. That problem is somewhat mitigated in our case given how most intelligence reports involve "key judgments" sections summarizing relevant aspects of uncertainty. The third section of our paper describes how we sought to replicate this kind of presentation when designing our experiments, but it would still be a mistake to interpret experimental results as externally valid representations of treatment effects. Instead, the main value of gathering this information is to understand whether respondents find our conceptions of confidence to be intuitively meaningful. If they do, this would suggest that existing approaches to assessing analytic confidence systematically sacrifice valuable information, even if our research design does not allow us to estimate exactly how valuable this information would be in any particular case.[4]

A third drawback with survey experiments is that respondents often devote limited attention to survey tasks (Berinsky, Margolis, & Sances, 2014). In our context, however, limited attention only raises the challenge of obtaining clear results. To the extent that analysts and decision-makers were able to engage with abstract elements of uncertainty in low-stakes survey experiments, this makes it *less* plausible to argue that professionals with greater experience, training, and incentives for effort could not also discriminate among our ideas.[5]

[3] Though the previous section highlighted examples of how individual conceptions of confidence shaped high-stakes decisions, we are unaware of a single case where national security officials discussed all three ideas together.

[4] Of course, we would be skeptical of *any* research design that claimed to offer such remarkable payoffs.

[5] Some surveys are also susceptible to social desirability bias, decreasing the proportion of respondents who are willing to offer controversial opinions (Blair, Imai, & Lyall, 2014). Yet because our hypotheses (described below) involve analyzing treatment effects rather than response means, this should not influence our analysis.

Finally, scholars must consider how survey respondents differ from populations of interest (Hafner-Burton, Hughes, & Victor, 2013). When studying how our three conceptions of confidence shape *decision-making*, we addressed this issue by recruiting an elite sample of 223 national security professionals enrolled in mandatory midcareer education at the National War College.[6] The National War College draws active-duty officers from all U.S. military services, along with professionals from civilian national security agencies (32% of respondents) and military officers from countries besides the United States (12% of respondents). Response rates at advanced military education programs are also high (ours was 88%). Our first survey experiment thus comprised a relatively broad cross-section of national security professionals with much less exposure to response bias than most other elite studies.

Our second study examined the extent to which political *analysts* who lacked specialized training could discriminate among different dimensions of analytic confidence. To conduct this study, we recruited 1,000 respondents via Amazon Mechanical Turk (AMT), which is an online platform where individuals complete surveys in exchange for compensation.[7] Since respondents recruited from AMT possess neither special interest nor special expertise in analyzing national security issues,[8] they are distinct from national security analysts or other professional assessors of uncertainty. In the context of our research, however, this should only increase the challenges of confirming that respondents could consistently disentangle our three conceptions of confidence. If we find that these individuals systematically grasp the distinctions among these ideas, then professional political analysts should be able to so even more effectively.

How National Security Professionals Interpret Analytic Confidence

Our first survey experiment employed a vignette-based design, presenting respondents with three national security decisions under uncertainty: a proposed hostage rescue mission with uncertainty about whether the hostages were being held in a particular location, a proposed drone strike with uncertainty about whether the target was a high-ranking terrorist, and a proposal to restrict air travel in response to a reported terrorist plot with uncertainty about whether this threat was genuine. The online supporting information provides the full text of each scenario.

Each vignette contained one paragraph describing the decision problem and one paragraph presenting assessments of uncertainty. The second paragraph began with a probability estimate, such as the chances that the target of the proposed drone strike was a high-ranking terrorist. We randomly assigned one of two values to these probability assessments, which differed by 15 percentage points. Then we briefly presented our three conceptions of confidence in random order. Each confidence assessment took one of two possible values. For example, here is the full text of the hostage rescue scenario:

> The U.S. military is searching for five American citizens held hostage by a rebel group overseas. They receive information suggesting that the hostages are being held in a rural compound. Analysts can tell that the compound is being used by the rebel group, but they have difficulty confirming that the hostages are present. Special forces officers expect that a raid on the compound will meet armed resistance. They do not want to

[6] We administered this survey between August 9 and 17, 2016. Mid-career programs like the National War College are required for U.S. military officers gaining promotion to the rank of colonel or commander.

[7] On the use of AMT for political science research, see Berinksy, Huber, & Lenz, 2012. We administered this survey on June 30, 2016. Respondents were 46% female and 80% White. Forty-seven percent had a college degree. Respondents were compensated $1.40 for finishing a survey that had a median completion time of 11 minutes. This rate of hourly compensation exceeded the federal minimum standard.

[8] On average, respondents reported spending less than one hour per day reading, watching, or discussing political news.

> put their soldiers in harm's way if the hostages are not present, but if they delay action too long, the rebels might move the hostages to a different location.
>
> After careful deliberation, a group of intelligence analysts assesses that there is a [60/75] percent chance that the hostages are being held inside this compound. The analysts explain that they have [little/a large amount of] reliable evidence on which to base their judgment; that there is [minimal/significant] disagreement among them about the chances that the hostages are present; and that they [do not believe their assessment would substantially change/believe their assessment could substantially change] if they continue to investigate the compound.

We designed this experiment to test three hypotheses. Our principal hypothesis was that each of our three conceptions of confidence would independently shape the manner in which respondents evaluated national security decisions. This is a demanding hypothesis. Reliability of available evidence, range of reasonable opinion, and responsiveness to new information are abstract ideas. If survey respondents reacted to these ideas simultaneously and independently, then this would indicate that national security professionals intuitively find that information to be meaningful when evaluating decisions under uncertainty.

Our theoretical framework also suggests that respondents should prioritize different conceptions of confidence when evaluating different aspects of national security decisions. For example, the article's first section explained how responsiveness to new information is most important for evaluating the trade-off between acting immediately and delaying a decision. By contrast, we argued that reliability of available evidence and range of reasonable opinion are primarily useful for evaluating a decision's political and ethical dimensions. Thus, our second hypothesis was that respondents would assign responsiveness to new information its greatest weight when evaluating the trade-off between action and delay, and our third hypothesis was that respondents would give the greatest weight to reliability of available evidence and range of reasonable opinion when evaluating a decision's political and ethical dimensions. Confirming that respondents prioritize different conceptions of confidence in this way would demonstrate note only that they paid attention to this information, but also that they engaged with these ideas in distinctive ways that match theoretical expectations.

> *H1*: All three conceptions of confidence will independently influence the way that respondents evaluate proposed decisions under uncertainty.
>
> *H2*: Responsiveness to new information will have its greatest effect in shaping respondents' perceptions about the trade-offs between acting immediately and gathering additional information.
>
> *H3a*: Reliability of available evidence will have its greatest effect in shaping respondents' perceptions about the political and ethical dimensions of decisions under uncertainty.
>
> *H3b*: Range of reasonable opinion will have its greatest effect in shaping respondents' perceptions about the political and ethical dimensions of decisions under uncertainty.

To test these hypotheses, we asked respondents to indicate levels of agreement (on a 7-point scale) with four statements after reading each scenario: "Decision makers should approve this proposal"; "Decision makers should gather more information before making a choice"; "It is ethically problematic to take action based on the information available in this scenario"; and "It would be difficult to defend this action to the public." The online supporting information presents descriptive statistics for each measure.

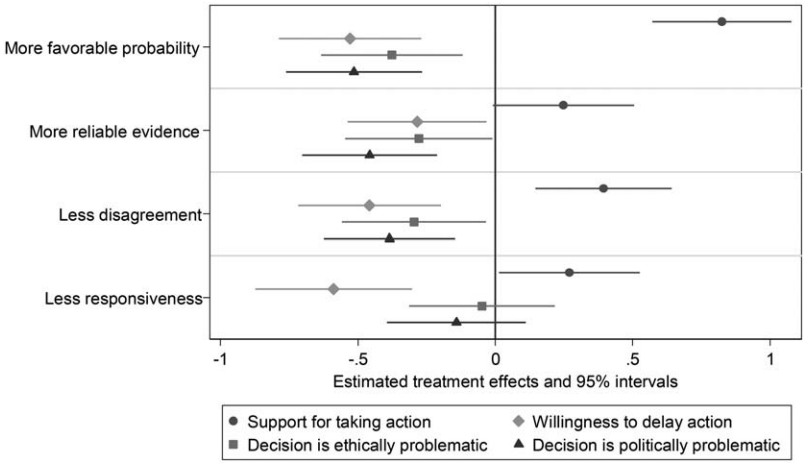


Figure 3. Responses to national security scenarios. Estimated treatment effects from OLS regressions characterizing respondents' reactions to national security decisions. Dependent variables measured on 7-point scales.

Results

We analyzed survey data using ordinary least squares regressions with fixed effects for each respondent and scenario.[9] (The online supporting information contains full results and alternative specifications.) Figure 3 summarizes our principal findings, showing how our four experimental treatments influenced each of the ways we asked respondents to evaluate decisions under uncertainty.

The data confirmed Hypothesis 1, as assessments of probability, reliability of available evidence, range of reasonable opinion, and responsiveness to new information all simultaneously influenced the manner in which respondents evaluated proposed courses of action. Each of these variables was a statistically significant ($p < 0.05$) predictor of at least two of the four response measures we elicited. On the central question of the extent to which respondents supported taking action in each case, assessments of probability, range of reasonable opinion, and responsiveness to new information each exerted effects that were statistically significant at the $p < 0.05$ level, while reliability of available evidence only narrowly missed that threshold ($p = 0.06$).

Our results also confirmed Hypothesis 2, showing that respondents gave responsiveness to new information its greatest weight when evaluating trade-offs between acting immediately and gathering additional information. The substantive impact of responsiveness to new information was twice as large for predicting willingness to delay action than for any of the other three measures we elicited from respondents.

Our experiment produced mixed support for Hypotheses 3a and 3b. As anticipated, we found that the reliability of available evidence shaped respondents' views most extensively when evaluating prospects for political controversy. Our results demonstrate that respondents saw range of reasonable opinion as having important political ramifications as well. Contrary to our expectations, however, we found that range of reasonable opinion had its greatest weight in shaping respondents' willingness to gather additional information.

To explore this pattern further, we analyzed how respondents combined information about range of reasonable opinion and responsiveness to new information when evaluating trade-offs between

[9] Mean (standard deviation) support for taking action was 4.9 (1.8) for the hostage scenario, 4.2 (2.0) for the drone strike scenario, and 2.7 (1.5) for the terrorism scenario. The online supporting information provides additional descriptive statistics.

Table 1. How Respondents Reacted to Range of Reasonable Opinion Versus Responsiveness to New Information

	High Responsiveness	Low Responsiveness	*Difference*
Substantial disagreement	5.32 (1.72)	5.01 (1.77)	0.31 ($p = 0.10$)
Little disagreement	5.19 (1.83)	4.31 (1.84)	0.88 ($p < 0.001$)
Difference	0.12 ($p = 0.53$)	0.70 ($p < 0.001$)	

Note. Mean support for delaying action (standard deviations in parentheses) based on assessments of range of reasonable opinion and responsiveness to new information; *p*-values reflect comparisons of means in two-way *t*-tests.

acting immediately and gathering additional information. Table 1 shows that respondents were most willing to delay action when analysts reported that their findings could change substantially in light of new information. In this context, range of reasonable opinion had no statistical impact on respondents' views. By contrast, when analysts reported that their findings were unlikely to change in response to new information, range of reasonable opinion explained substantial variation in respondents' willingness to delay. These patterns suggest that respondents implicitly applied a two-step decision rule when considering the benefits of gathering additional information. The primary conception of confidence that respondents examined when considering this trade-off was the extent to which delaying action would produce informational gains. Even when respondents were skeptical about the informational value of delay, however, they were still hesitant to act as long as analysts disagreed over how to assess uncertainty.

These experimental results demonstrate that our three conceptions of confidence capture independent elements of high-stakes decision-making, elements that matter not merely in principle but also in practice. Though distinctions among our three conceptions of confidence are abstract, and though we gave respondents no special instructions about how to interpret this information, we found that these attributes influenced the way that national security professionals intuitively evaluated decisions under uncertainty. Not only did our respondents react to these cues, but they also prioritized different elements of analytic confidence when assessing specific elements of decision-making in ways that either matched our theoretical expectations or suggested otherwise sensible approaches to managing uncertainty.

How Inexperienced Analysts Assess Analytic Confidence

Even if our three conceptions of confidence independently shape the way that national security officials evaluate high-stakes decisions, these ideas would still offer little practical value if analysts could not reliably distinguish among them. To investigate whether political analysts can intuitively disentangle the dimensions of analytic confidence, we administered a second survey experiment to 1,000 respondents via Amazon Mechanical Turk (AMT). As our second section described, this research design provides a tough test of conceptual validity. To the extent that inexperienced analysts can consistently discriminate among our three conceptions of confidence, we should expect professional analysts with greater training and expertise to make even greater use of these ideas.

We asked each respondent to answer 10 questions, randomly chosen from a broader list of 20. Eight questions pertained to U.S. domestic politics, such as "What are the chances that Hillary Clinton will win the 2016 U.S. presidential election?" Eight questions pertained to foreign policy issues, such as "What are the chances that Britain will formally exit the European Union by the end of 2016?" The remaining questions dealt with issues that involved both foreign and domestic politics, such as "What are the chances that the United States will accept more than 7,500 Syrian refugees by the end of 2016?" Figure 4 summarizes these questions, and the online supporting information provides complete wordings.

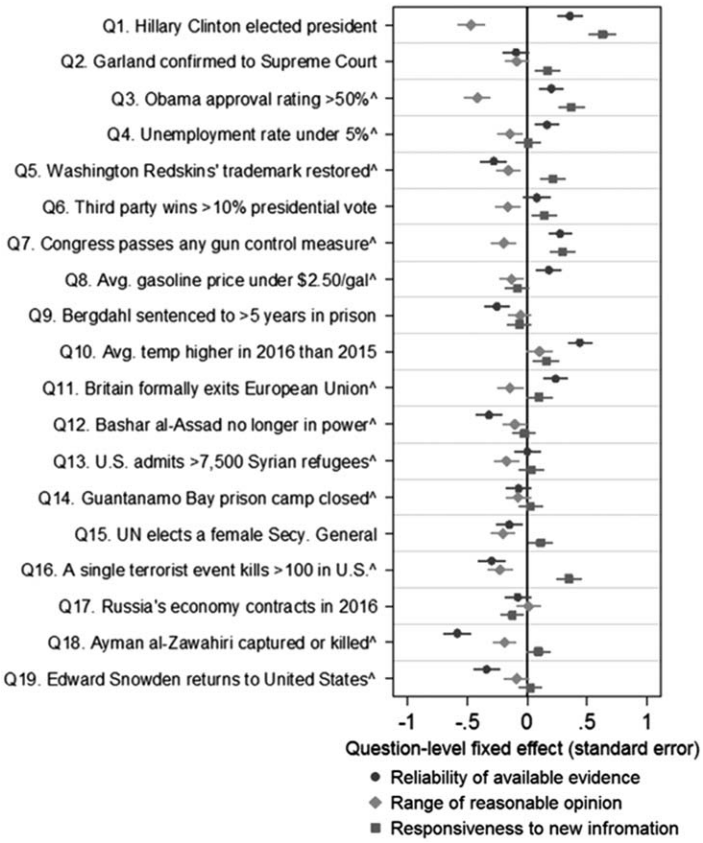


Figure 4. Exploring independent variation in conceptions of confidence across questions. Mean estimates (with standard errors) for how respondents' assessments of each conception of confidence deviated from the norm on each question, controlling for probability assessments, additional conceptions of confidence, and respondent demographics. We withheld Question 20 as a baseline, which asked about the chances that more than 10 U.S. soldiers would be killed in Iraq in 2016; "^" designates that the question specified the time frame was by/at the end of 2016.

For each question, we asked respondents to provide a probability estimate. Then we asked respondents to indicate the extent to which they agreed with each of the following statements: "I could defend this estimate with a substantial amount of reliable information"; "Reasonable people could give substantially different answers to this question"; and "My answer could change substantially if I studied this subject further." We elicited agreement with these statements on 5-point scales, as this is the level of precision with which intelligence analysts generally assign confidence levels (low, low-to-moderate, moderate, moderate-to-high, high). After gathering these data, we ordered each scale so that larger values indicated greater confidence.

Identifying Independent Variation

Respondents assigned the same value to all three conceptions of confidence just 17% of the time. For 25% of observations, the confidence levels that respondents attached to their forecasts differed by at least three points out of five across measures. Overall correlations among these measures ranged from 0.27 to 0.53, displaying substantial independent variation in the ways respondents assessed our three conceptions of confidence. The online supporting information analyzes these descriptive statistics in greater detail.

This independent variation could nevertheless reflect arbitrary responses to survey questions as opposed to systematic distinctions among abstract ideas. If the relationship between our three measures tended to be stable on average, and if the remaining variation was simply random noise, then there would be little practical value in asking analysts to distinguish among these concepts. The key empirical issue is thus whether respondents' confidence levels varied across measures in a manner that was both independent and consistent.

To evaluate this claim, we analyzed variation in how respondents assessed each conception of confidence while controlling for the degree of certainty that respondents assigned to each forecast (defined as the absolute value of the difference between each probability estimate and 50%), as well as how respondents assessed the other two kinds of confidence when making each forecast. After controlling for these measures, we added dummy variables for each of our survey questions.[10] The coefficients on these dummy variables thus indicate the extent to which respondents' assessments of confidence varied across questions in ways that we cannot explain as a function of how they assessed other elements of uncertainty. If these question-level fixed effects prove to be statistically significant, this would confirm that analysts consistently discriminated among our three conceptions of confidence. In other words, this would show that the independent variation among these conceptions of confidence reflected systematic distinctions as opposed to arbitrary detail.[11]

Results

Figure 4 presents results, estimated using ordinary least squares regression with respondent fixed effects and standard errors clustered by respondent. (Again, see the online supporting information for full analysis and additional specifications.) There was only one question on which we found no statistically significant difference between the way that respondents assessed each conception of confidence and what we would have predicted given the other covariates in our model. Sixty-five percent of these question-level fixed effects are statistically significant at the $p < 0.05$ level, and 40% of them are statistically significant at the $p < 0.001$ level.

To illustrate our findings, Figure 4 shows that when analysts estimated the chances that al-Qaeda's leader would be killed or captured by the end of the year (question 19), they accepted that this judgment was based on an unusually limited amount of reliable information, even after controlling for the way that those respondents evaluated probability and the other two kinds of confidence in describing their beliefs. When respondents estimated the chances that Barack Obama's approval rating would be above 50% by the end of 2016 (question 3), they described this judgment as being unusually open to reasonable disagreement. When respondents estimated the chances that the U.S. Congress would pass gun control legislation by the end of 2016 (question 7), they believed that their views were particularly unlikely to change in response to new information. For 6 of 20 questions, we observed independent and statistically significant departures from the mean for *all three* conceptions of confidence at once. For example, when respondents estimated the chances that Hillary Clinton would be elected president (question 1), they generally reported that their beliefs were based on unusually large amounts of reliable evidence and that those beliefs were unusually unlikely to change in response to new information, but that reasonable people were also unusually likely to disagree when offering their views on this issue.

[10] We withheld the twentieth question as a baseline.

[11] Our analyses also contain a battery of demographic controls. We included information on college education (dummy variable), age, and political engagement (the number of hours per week respondents reported watching, reading, or discussing political news) on the grounds that older and/or more informed respondents might assign higher confidence to their judgments. We included information on gender, race, and political ideology given research suggesting that men (Johnson et al., 2006) and conservatives (Ortoleva & Snowberg, 2015b) are especially prone to overconfidence, and that Whites differ from other groups when assessing risk (Finucane, Slovic, Mertz, Flynn, & Satterfield, 2010). See the online supporting information for details.

These results demonstrate that respondents consistently discriminated among our three conceptions of confidence when making political forecasts. Though these respondents had no particular experience analyzing politics, and though we provided no special instructions for disentangling abstract elements of assessing uncertainty, we found that they intuitively grasped the differences among these abstract ideas. Thus, while our first survey experiment found that these distinctions matter to the ways that national security professionals evaluate decisions under uncertainty, our second survey shows that even inexperienced assessors can draw these distinctions in practice.

Summary and Directions for Further Research

In 2002, the United States Intelligence Community released a National Intelligence Estimate, judging with "high confidence" that Iraq was "continuing, and in some areas expanding, its chemical, biological, nuclear and missile programs." What did it mean to make this judgment with high confidence? One interpretation is that analysts thought it was highly likely that Iraq was pursuing weapons of mass destruction. Perhaps this statement was 90% likely to be true: a "slam dunk" as this statement was described in public debate. Or maybe analysts thought the chances that Iraq was pursuing weapons of mass destruction were more like 60%, and they used the term "high confidence" to indicate that most analysts believed this judgment was likely to be correct. While the difference between these interpretations could have had markedly different implications for debates about invading Iraq, there was little discussion at the time, either in government or in the public sphere, about what exactly it meant to assess uncertainty with "high confidence" (Jervis, 2010, chap. 3).

Over the last decade, intelligence agencies in the United States and in several other countries have undertaken substantial efforts to reduce confusion when communicating uncertainty (Ho, Budescu, Dhami, & Mandel, 2015; Wheaton, 2012). Yet scholars and practitioners still lack conceptual foundations for describing what analytic confidence means in principle, and there remain open empirical questions about the extent to which analysts and decision-makers can intuitively grasp these ideas in practice. To address these issues, we argued for disaggregating "analytic confidence" into three distinct attributes: reliability of available evidence, range of reasonable opinion, and responsiveness to new information. We showed that these attributes can vary independently and explained why they hold different implications for high-stakes decision-making. Our survey experiments demonstrated that these ideas shape the way that national security professionals evaluate decisions under uncertainty and that even inexperienced analysts can consistently discriminate among these abstract ideas. Though our analysis focused primarily on the challenges of assessing uncertainty in national security, we explained how similar issues surround virtually any other area of political decision-making. To our knowledge, this article represents the first attempt to build theoretical and empirical foundations for understanding analytic confidence in this way.

Further research can extend our analysis in at least three directions. One clear avenue for additional research is to explore how different conceptions of confidence covary across issues. It is particularly important to determine the conditions under which our three conceptions of confidence are most likely to diverge, as these are the cases where separating conceptions of confidence would likely provide the most practical value for shaping decisions under uncertainty.

A second extension of this research would employ within-case process tracing to explore how decision-makers respond to different conceptions of confidence when making high-stakes choices. Since survey experiments cannot capture the dense mixture of cues and pressures that real high-stakes decision entail, we have limited basis for making clear predictions about exactly where these conceptions of confidence have their greatest impact. Further work on this subject—perhaps delving into the cases we described in the article's first section—would hone scholars' ability to describe how different components of uncertainty shape major decisions.

A third potential extension of this research program would explore the psychological mechanisms by which decision-makers interpret and combine our three conceptions of confidence. Though many scholars question the extent to which individuals can grapple with the nuances of assessing uncertainty, our survey experiments showed how both elite and nonelite respondents could consistently separate probability from confidence and that they could reliably disentangle confidence into three distinct components. Our first survey experiment further suggested that decision-makers interpreted responsiveness to new information and range of reasonable opinion using a sensible, multistep process. Altogether, these findings offer encouraging implications regarding individuals' capabilities to engage with abstract elements of uncertainty, inviting further experimental investigation of what these thought processes entail.

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Homeland Security (DHS) through the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) at the University of Southern California (USC) under award number 2010-ST-061-RE0001. Any opinions, findings, conclusions, or recommendations in this document are those of the authors and do not necessarily reflect views of DHS, USC, or CREATE. Geoffrey Berlin, Mark Bucknam, Stephen Dyson, Thomas Fingar, Paul Heer, Yusaku Horiuchi, Dean Lacy, Mark Lowenthal, Brendan Nyhan, Gregory Treverton, Kristan Wheaton, and Laura Yost provided helpful comments on previous drafts, as did seminars at the National Intelligence Council, the Department of Homeland Security's Science and Technology Directorate, the 84th Military Operations Research Society Symposium, the Ridge College of Intelligence Studies and Applied Sciences at Mercyhurst University, and the 2017 annual meeting of the International Studies Association. Support through the ANR - Labex IAST is gratefully acknowledged. Freya Jamison and Ben Rutan provided outstanding research assistance. We are especially grateful to the 223 national security officials who volunteered their time to participate in our survey experiments. Correspondence concerning this article should be addressed to Jeffrey A. Friedman, Department of Government, Dartmouth College, Silsby Hall HB 6108, Hanover, NH 03766. E-mail: jeffrey.a.friedman@dartmouth.edu

REFERENCES

Achen, C., & Bartels, L. (2016). *Democracy for realists*. Princeton, NJ: Princeton University Press.

Al-Najjar, N., & Weinstein, J. (2009). The ambiguity aversion literature. *Economics and Philosophy*, *25*, 249–284.

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid? *American Political Science Review*, *104*, 226–242.

Berinsky, A., Huber, G., & Lenz, G. (2012). Evaluating online labor markets for experimental research. *Political Analysis*, *20*, 351–368.

Berinsky, A., Margolis, M., & Sances, M. (2014). Separating the shirkers from the workers? *American Journal of Political Science*, *58*, 739–753.

Betts, R. (2000). Is strategy an illusion? *International Security*, *25*, 5–50.

Betts, R. (2007). *Enemies of intelligence*. New York, NY: Columbia University Press.

Blair, G., Imai, K., & Lyall, J. (2014). Comparing list and endorsement experiments. *American Journal of Political Science*, *58*, 1043–1063.

Doyle, M. (2008). *Striking first: Preemption and prevention in international conflict*. Princeton, NJ: Princeton University Press.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*, 643–669.

Fingar, T. (2011). *Reducing uncertainty: Intelligence analysis and national security*. Stanford, CA: Stanford University Press.

Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2010). Gender, race, and perceived risk: The 'white male' effect. *Health, Risk & Society*, 2, 159–172.

Friedman, J., & Zeckhauser, R. (2015). Handling and mishandling estimative probability. *Intelligence and National Security*, *30*, 77–99.

Gigerenzer, G. (2008). *Rationality for mortals*. New York, NY: Oxford University Press.

Hafner-Burton, E., Hughes, A., & Victor, D. (2013). The cognitive revolution and the political psychology of elite decision making. *Perspectives on Politics*, *11*, 368–386.

Hayden, M. (2016). *Playing to the edge*. New York, NY: Penguin.

Ho, E., Budescu, D., Dhami, M., & Mandel, D. (2015). Improving the communication of uncertainty in climate change and intelligence analysis. *Behavioral Science and Policy*, *1*, 43–55.

Hyde, S. (2015). Experiments in international relations. *Annual Review of Political Science*, *18*, 403–424.

Jervis, R. (1997). *Systems effects*. Princeton, NJ: Princeton University Press.

Jervis, R. (2006). Reports, politics, and intelligence failures. *Journal of Strategic Studies*, *29*, 3–52.

Jervis, R. (2010). *Why intelligence fails*. Ithaca, NY: Cornell University Press.

Johnson, D. (2004). *Overconfidence in war*. Cambridge, MA: Harvard University Press.

Johnson, D., McDermott, R., Barrett, E., Cowden, J., Wrangham, R., McIntyre, M., & Rosen, S. (2006). Overconfidence in wargames. *Proceedings of the Royal Society B*, *273*, 2513–2520.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: FSG.

Kent, S. (1964). Words of estimative probability. *Studies in Intelligence*, *8*, 49–65.

Loewenstein, G., Hsee, C., Weber, E., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*, 267–286.

Manksi, C. (2013). *Public policy in an uncertain world*. Cambridge, MA: Harvard University Press.

Mastrandrea, M., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., … Zwiers, F. W. (2010). *Guidance note for lead authors*. Jasper Ridge, CA: IPCC.

McAuliff, C. (1982). Burdens of proof. *Vanderbilt Law Review*, *35*, 1293–1335.

Moore, D., & Healy, P. (2008). The trouble with overconfidence. *Psychological Review*, *115*, 502–517.

Morell, M. (2015). *The great war of our time*. New York, NY: Crown.

Neustadt, R., & May, E. (1986). *Thinking in time*. New York, NY: Free Press.

Nye, J., Jr. (1994). Peering into the future. *Foreign Affairs*, *73*, 82–93.

Ortoleva, P., & Snowberg, E. (2015a). Overconfidence in political behavior. *American Economic Review*, *105*, 504–535.

Ortoleva, P., & Snowberg, E. (2015b). Are conservatives overconfident? *European Journal of Political Economy*, *40B*, 333–344.

Rathbun, B. (2007). Uncertain about uncertainty: Understanding the multiple meanings of a crucial concept in international relations theory. *International Studies Quarterly*, *51*, 533–557.

Slovic, P. (2010). *The feeling of risk*. London, United Kingdom: Earthscan.

Tetlock, P. (2005). *Expert political judgment*. Princeton, NJ: Princeton University Press.

U.S. Senate. (2004). *Report on the U.S. intelligence community's prewar intelligence assessments of Iraq*. Washington, DC: Select Committee on Intelligence.

Vertzberger, Y. (1995). Rethinking and reconceptualizing risk in foreign policy decision-making. *Political Psychology*, *16*, 347–380.

Wheaton, K. (2012). The revolution begins on page five: The changing nature of NIEs. *International Journal of Intelligence and CounterIntelligence*, *25*, 330–349.

Zaller, J. (1992). *The nature and origins of mass opinion*. New York, NY: Cambridge University Press.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website:

Full Text of Vignettes and Experimental Manipulations

Distribution of Responses to Scenarios

Full Results for Survey Experiment 1