The Use of Alternative Reasons in Probabilistic Judgments

Burcu Gürçay & Jonathan Baron
University of Pennsylvania

Goals of the Study

- Create an online training module to teach people AOT behavior

- Test the new behavioral AOT measure (AOT-Reasons)

Myside bias

- People tend to favor possibilities that are already strong during search and inference.

- People ignore or dismiss evidence that conflicts with their favored possibilities.

Measuring individual differences in actively open-minded thinking

- Belief measures (Stanovich & West; Baron)

- Behavioral measures

Original AOT scale: Beliefs about thinking

Response scale: Strongly disagree….. Strongly agree (5 points)

- Allowing oneself to be convinced by an opposing argument is a sign of good character.

- People should take into consideration evidence that goes against their beliefs.

- People should revise their beliefs in response to new information or evidence.

- Changing your mind is a sign of weakness. (-)

- Intuition is the best guide in making decisions. (-)

- It is important to persevere in your beliefs even when evidence is brought to bear against them. (-)

- One should disregard evidence that conflicts one's established beliefs. (-)

Problems with belief measures

- Social desirability effect

- Lack of self-control

Behavioral measures of AOT

- Measuring actual thinking rather than beliefs about how we should think

- Argument Evaluation Test (AET) (Stanovich & West, 1997 & 1998)

- Other similar methods to AET (e.g., Baron 1995 and 2009)

Problems with existing behavioral measures

- They require a lot of items to derive a meaningful AOT measure

- They need to be scored by experts or experimenters

- The items used need to appeal to a wider population (and be age appropriate in the case of younger populations)

PART I:
Measuring Individual Differences in Actively Open-Minded Thinking

N	131
Median$_{age}$	45
Age Range	18 to 76
% Female	66.4
Number of Questions	20
Question Type	3 option MC

- Within-Subject Study

- Control and training conditions had 60 and 51 subjects, respectively.

- All subjects answered the questions twice.

- Training condition subjects had to go through an hour long training session before answering the questions for the second time.

Which is the movie that has the most recent release date?
Answer A. V for Vendetta
Answer B. The Matrix
Answer C. Se7en

What answer would you give without thinking about it?
○ A   ● B   ○ C

Now think about it and list the reasons you think of.

Brad Pitt is quite young in Se7en.

Which answer is this argument about?
○ for A   ○ for B   ○ for C   ○ against A   ○ against B   ● against C   ○ other*

What is the probability that answer A (V for Vendetta) is correct (in %)?
○0 ○1 ○5 ○10 ○20 ○30 ○40 ○50 ○60 ○70 ○80 ○90 ○95 ○99 ○100

What is the probability that answer B (The Matrix) is correct (in %)?
○0 ○1 ○5 ○10 ○20 ○30 ○40 ○50 ○60 ○70 ○80 ○90 ○95 ○99 ○100

What is the probability that answer C (Se7en) is correct (in %)?
○0 ○1 ○5 ○10 ○20 ○30 ○40 ○50 ○60 ○70 ○80 ○90 ○95 ○99 ○100
If you guess, you have a 33% chance of being correct.

[Click here when you have answered all questions.]

* The 'other' response means that the reason does not distiguish the choices from each other.

Please write any comments on this page here (up to 255 characters):
[   ]

Individual Difference Measures in AOT

AOT-Beliefs: AOT Scale Scores

8. People should search actively for reasons why their beliefs might be wrong.

9. When we are faced with a new question, the first answer that occurs to us is usually the best. (-)

10. When faced with a new question, we should consider more than one possible answer before reaching a conclusion.

11. When faced with a new question, we should look for reasons why our first answer might be wrong, before deciding on an answer.

	α
ROUND 1	0.81
ROUND 2	0.86

Individual Difference Measures in AOT

AOT-Reasons: AOT Scores from Subjects' Reasons

Ss received 1 point for giving a reason that is for an option other than preferred option or a reason that is against their preferred option.

No point was given for reasons that were for the preferred option.

There were no statistically significant correlations between AOT-Beliefs and AOT-Reasons measures.

AOT Measures and Accuracy

<u>Brier Scores</u>

We used Brier Scores to measure accuracy.

The squared difference between

- the predicted probability assigned to the possible outcomes

- and the actual outcome.

Lower scores indicate higher accuracy.

AOT Measures and Accuracy

- Are higher mean AOT-Beliefs scores associated with higher mean accuracy (lower mean Brier scores)?

- Are higher mean AOT-Reasons scores associated with higher mean accuracy (lower mean Brier scores)?

ACCURACY SCORES

	Round 1	Round 2
AOT-Beliefs	-0.31**	-0.21**
AOT-Reasons	-0.05	-0.12

** indicates statistically significant result in the predicted direction in a one-tailed test.

Effects of AOT Measures on Accuracy

Linear mixed effect models with subjects and items as random effects

Accuracy scores ~ AOT measures

Dependent Variable	Independent Variable	Round 1	Round 2
Brier Scores	AOT-Reasons	(+)	(+)
Brier Scores	AOT-Beliefs	+	+

+   statistically significant effect in the expected direction

(+) statistically non-significant effect in the expected direction
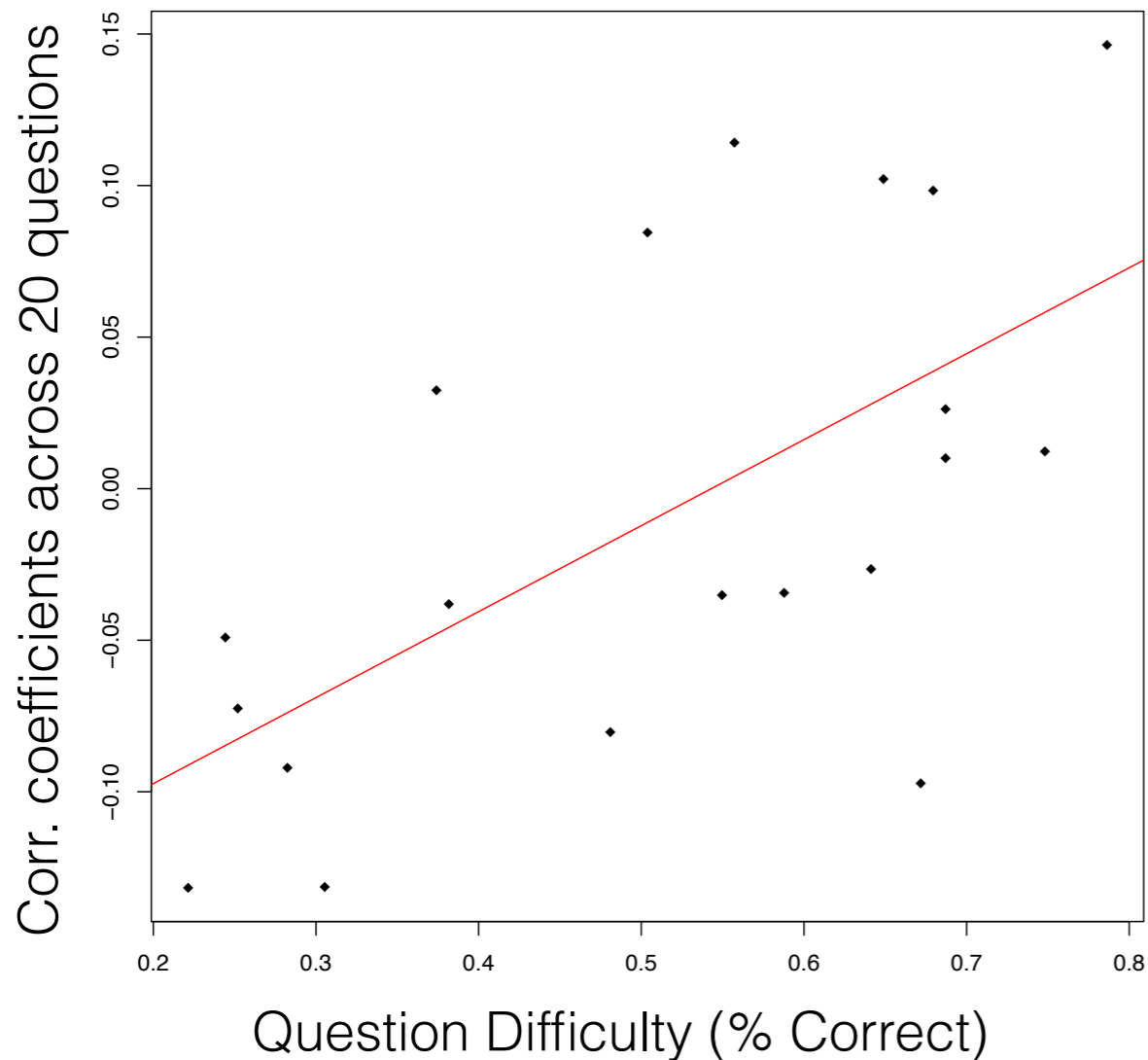
AOT Measures and Accuracy

- Haran, Ritov, & Mellers (2013): Higher scores on AOT Scale were associated with lower overconfidence.

- Brier scores use subjects' probability judgments to calculate accuracy.

- AOT might benefit subjects by lowering their extreme confidence judgments.

- This mechanism is most beneficial in cases where the probability of wrong answer is high (i.e., difficult questions).
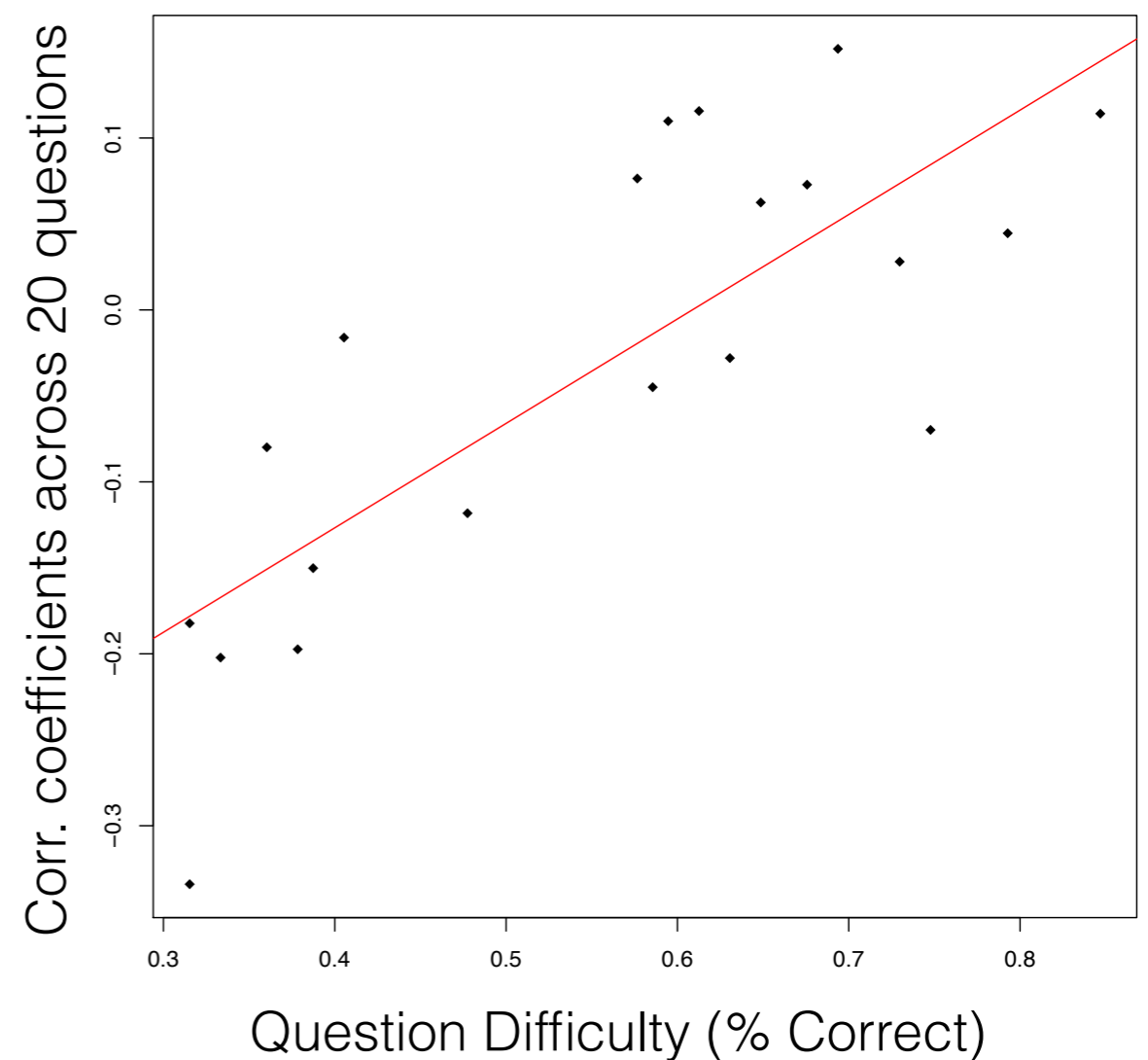
AOT-Reasons and Brier Scores

How does the correlation between AOT-Reasons and Brier Scores change as the question difficulty varies?

Round 1

$r = -0.31$

Round 2

$r = -0.58*$

Effects of AOT Measures on Accuracy

Is the usefulness of considering alternative reasons dependent on knowing the correct answer or not?

Dependent Variable	Independent Variables	Round 1	Round 2
Brier Scores	AOT-Reasons	(+)	+
	Correct	+	+
	AOT-Reasons* Correct	+	+

+   statistically significant effect in the expected direction

(+)   statistically non-significant effect in the expected direction

Effects of AOT Measures on Accuracy

Is the usefulness of considering alternative reasons dependent on knowing the correct answer or not?

Data	Dependent Variable	Independent Variable	Round 1	Round 2
correct answers	Brier Scores	AOT-Reasons	+	+
incorrect answers	Brier Scores	AOT-Reasons	+	+

+   statistically significant effect in the expected direction

(+)   statistically non-significant effect in the expected direction

Overconfidence Measures

We assessed the relations between overconfidence and AOT measures with two different overconfidence measures:

- **Over1** compared subjects' probability judgments for their preferred answers to the percentage of questions they answered correctly.

- **Over2** utilized all three probability judgments, used Brier scores, and was also loosely analogous to the first overconfidence measure.

For both measures, positive numbers indicate overconfidence.

	Mean$_{Over1}$	Mean$_{Over2}$
ROUND 1	0.11	0.19
ROUND 2	0.08	0.17

22

Overconfidence and AOT Measures

The correlation coefficients (*r*) between AOT and overconfidence measures are shown below.

	VARIABLES	Round 1	Round 2
Over1	AOT-Reasons	-0.10	-0.18**
Over1	AOT-Beliefs	-0.28**	-0.21**
Over2	AOT-Reasons	-0.13	-0.21**
Over2	AOT-Beliefs	-0.28**	-0.22**

** indicates statistically significant result in the expected direction in a one-tailed test.

Effects of AOT Measures on Overconfidence

Linear mixed effect models with subjects and items as random effects

Overconfidence scores ~ AOT measures

Data	Dependent Variable	Independent Variable	Round 1	Round 2
all	Over2	AOT-Reasons	+	+
all	Over2	AOT-Beliefs	+	+
all	$Prob_{preferred}$	AOT-Reasons	+	+
all	$Prob_{preferred}$	AOT-Beliefs	+	+

  +   statistically significant effect in the expected direction
(+) statistically non-significant effect in the expected direction

PART II:
Teaching Actively Open-Minded Thinking

Teaching AOT

- Previous studies showed that AOT can be taught through instruction
  (Selz, 1935; Perkins et al., 1986; Baron et al., 1985; Graumlich & Baron, 1991)

- There are no AOT training studies with adults

- The existing AOT training modules take weeks or months and require an instructor

Overview

- We trained half of our subjects in AOT between the two rounds of questions.

- We looked at the effects of this training on:

  - the number of alternative reasons subjects wrote

  - subjects' accuracy scores

  - subjects' confidence

Online AOT training

- Introduction to the training

- Introduction to the concept of AOT and related vocabulary

- Introduction to the concept of myside bias and how to avoid it

- An exercise that gave subjects the opportunity to apply what they have learned during the training
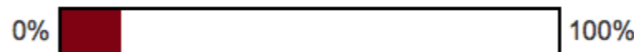
Online AOT training

Actively Open-Minded Thinking

We said earlier that *good thinking* is actively open-minded thinking. What do we mean by that?

Actively open-minded thinking consists of:
- Search for possibilities and evidence that is thorough enough for the task at hand
- Fair judgment of other possibilities and evidence other than the ones we initially favor
- Confidence that is assigned appropriately based on the quantity and the quality of thinking done.

AOT helps us correct mistakes before we make them. It also helps to make good judgments about probabilities. With AOT, we believe more strongly in true conclusions and less strongly in false ones.

0% [■■   ] 100%

NEXT

Online AOT training

Example Exercise Questions:

How should government agencies handle the manufacturing and distribution of nutritional supplements and vitamins?

Which of the following countries has the most forested area in square miles? (Answer options: Brazil, Canada, Russia)

Take a look at your reasoning again. Which categories did your reasoning fall in? (Select all that apply.)

☐ **Failed to look hard enough for counter-evidence** (I now see that there are arguments I didn't think of)

☐ **Looked for counter-evidence but didn't find any** (This is not an example of myside bias because you tried)

☐ **Under-weighed counter-evidence** (I thought of evidence against my answer but did not take it seriously enough)

☐ **Over-weighed evidence** (I took evidence for my answer too seriously)

☐ **Knew it** (I already knew the answer so I did NOT have to think)

☐ **Good AOT** (I thought of counter-evidence and used it appropriately)

Comparisons Between Training and Control Conditions

Hypotheses Tested	Results
Subjects in the training condition improved their AOT-Reasons scores more than those in the control condition.	+
Subjects in the training condition improved their AOT-Beliefs scores more than those in the control condition.	+
When subjects did not know the correct answer, those in the training condition improved their Brier scores more than those in the control condition.	+
When subjects knew the correct answer, those in the experimental condition improved their Brier scores more than those in the control condition.	(−)

+   statistically significant result in the expected direction
(+)   statistically non-significant result in the expected direction
—   statistically significant result in the opposite direction
(—) statistically non-significant result in the opposite direction

Comparisons Between Training and Control Conditions

Hypotheses Tested	Result
Subjects in the training condition improved their Brier scores more than those in the control condition.	(+)
Using Over1 as the overconfidence measure, subjects' overconfidence diminished from Round 1 to Round 2 more than those in the control condition.	+
Using Over2 as the overconfidence measure, subjects' overconfidence diminished from Round 1 to Round 2 more than those in the control condition.	(+)

+   statistically significant result in the expected direction
(+) statistically non-significant result in the expected direction

Discussion

- Belief vs. Behavioral Measures of AOT

- AOT and Accuracy

- AOT and Overconfidence

- Decreasing Myside Bias via Training

Thank you

Over2

3 probabilities

Ss can be overconfident in giving a prob that's too high to the wrong answer or one that's too low for the correct answer.

We wanted to calculate their overconfidence such that we would consider all 3 probabilities.

We still used the idea that confidence is the expected probability of success, and came up with EBS.

Let x, y, and z be the probabilities the subject assigns to each option on a three-choice question:

$$EBS = x(1-x)^2 + (1-x)x^2 + y(1-y)^2 + (1-y)y^2 + z(1-z)^2 + (1-z)z^2$$

Over2 = Brier scores – Expected Brier scores

This score is loosely analogous to Over1.

Over1 vs Over2

- binary task, p = S's prob judgment for an answer being correct

- EBS = $p(1-p)^2+(1-p)p^2 + p(1-p)^2+(1-p)p^2 = 2p(1-p)$

- If p is true, then BS = $(1-p)^2 + (1-p)^2 = 2(1-p)^2$

- If p is false, then BS = $(1-(1-p))^2+ (1-(1-p))^2 = 2p^2$

- Then EBS = $2p(1-p)^2+2(1-p)p^2 = 2p(1-p)$

- We can ignore the 2 coef. bc it is due to symmetry so

- EBS = $p(1-p)$

- Next, assume that Q is the actual proportion of cases like this that are correct, then BS = $2Q(1-p)^2+2(1-Q)p^2$

- Drop the 2 again, which gives us $(1-2p)(Q-p)$

- The corresponding ordinary measure is Q-p.

- We can write 1-2p as 2(0.5-p), and it coms fro the fact that prob below 0.5 should never get assigned to the prob of an event being true.

- this means that the function is relatively flat near p=0.5 compared to the usual overconfidence measure of p=Q.