

Critical Thinking Innovation

JAKE CHANDLER

Confirmation bias & overconfidence

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

Introduction

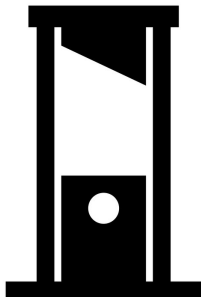
- Previously: errors *specific* to probabilistic thinking
- This week: a famous issue for *both* probabilistic *and* deductive thinking, **confirmation bias**

“perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”. (Evans, 1989)

- It is likely connected to a second problem: **overconfidence**
- We'll discuss
 - What the evidence is for both mistakes
 - Why we might have evolved to make them
 - How we can try to avoid them

SOME EXPERIMENTS

The death penalty experiment



The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment

- The experiment (Lord *et al* 1979):
 - Subjects were divided into two groups, according to prior views on the deterrence effect of capital punishment: opponents vs supporters
 - Both groups were presented with detailed discussions of two studies: one suggesting effective deterrence, the other opposing it.
 - Subjects were asked to:
 - rate the extent to which each study was well conducted and convincing
 - indicate the way in which the studies had influenced their opinion

The death penalty experiment (ctd)

- The results:
 - People exhibited significantly **biased quality assessments** in favour of the studies that confirmed their initial opinions
 - People's opinions became increasingly **polarised**, in spite of having view the very same evidence
- Moral:

We tend to view arguments supporting our own views more favourably than we do arguments that undermine those views

The death penalty experiment (ctd)

- The results:

- People exhibited significantly **biased quality assessments** in favour of the studies that confirmed their initial opinions
- People's opinions became increasingly **polarised**, in spite of having view the very same evidence

- Moral:

We tend to view arguments supporting our own views more favourably than we do arguments that undermine those views

The death penalty experiment (ctd)

- The results:
 - People exhibited significantly **biased quality assessments** in favour of the studies that confirmed their initial opinions
 - People's opinions became increasingly **polarised**, in spite of having view the very same evidence
- Moral:

We tend to view arguments supporting our own views more favourably than we do arguments that undermine those views

The death penalty experiment (ctd)

- The results:
 - People exhibited significantly **biased quality assessments** in favour of the studies that confirmed their initial opinions
 - People's opinions became increasingly **polarised**, in spite of having view the very same evidence

- Moral:

We tend to view arguments supporting our own views more favourably than we do arguments that undermine those views

The death penalty experiment (ctd)

- The results:
 - People exhibited significantly **biased quality assessments** in favour of the studies that confirmed their initial opinions
 - People's opinions became increasingly **polarised**, in spite of having view the very same evidence
- Moral:

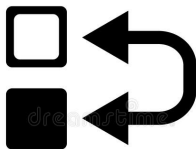
We tend to view arguments supporting our own views more favourably than we do arguments that undermine those views

The death penalty experiment (ctd)

- The results:
 - People exhibited significantly **biased quality assessments** in favour of the studies that confirmed their initial opinions
 - People's opinions became increasingly **polarised**, in spite of having view the very same evidence
- Moral:

We tend to view arguments supporting our own views more favourably than we do arguments that undermine those views

The answer inversion experiment



The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment

- The experiment (Hall *et al* 2012):
 - Subjects were asked to rate their degree of agreement on some moral claims, such as:

“Even if an action might harm the innocent, it can still be morally permissible to perform it. ”
 - They were then asked to justify their views
 - Plot twist: among the views presented to the subjects as their own were statements that contradicted their original claims, such as

“If an action might harm the innocent, it cannot be morally permissible to perform it. ”

The answer inversion experiment (ctd)

- The results:
 - Over two thirds of the subjects failed to notice the inversion (**choice blindness**: inability to recall a choice even shortly after having made it)
 - Over half of the participants **argued unequivocally for the opposite** of their original attitude in at least one instance
- Moral:

We are capable of readily providing arguments for views that we take to be our own even when we actually think they are false

The answer inversion experiment (ctd)

- The results:

- Over two thirds of the subjects failed to notice the inversion (**choice blindness**: inability to recall a choice even shortly after having made it)
- Over half of the participants **argued unequivocally for the opposite** of their original attitude in at least one instance

- Moral:

We are capable of readily providing arguments for views that we take to be our own even when we actually think they are false

The answer inversion experiment (ctd)

- The results:
 - Over two thirds of the subjects failed to notice the inversion (**choice blindness**: inability to recall a choice even shortly after having made it)
 - Over half of the participants **argued unequivocally for the opposite** of their original attitude in at least one instance
- Moral:

We are capable of readily providing arguments for views that we take to be our own even when we actually think they are false

The answer inversion experiment (ctd)

- The results:
 - Over two thirds of the subjects failed to notice the inversion (**choice blindness**: inability to recall a choice even shortly after having made it)
 - Over half of the participants **argued unequivocally for the opposite** of their original attitude in at least one instance

- Moral:

We are capable of readily providing arguments for views that we take to be our own even when we actually think they are false

The answer inversion experiment (ctd)

- The results:
 - Over two thirds of the subjects failed to notice the inversion (**choice blindness**: inability to recall a choice even shortly after having made it)
 - Over half of the participants **argued unequivocally for the opposite** of their original attitude in at least one instance
- Moral:

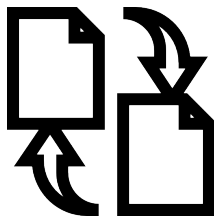
We are capable of readily providing arguments for views that we take to be our own even when we actually think they are false

The answer inversion experiment (ctd)

- The results:
 - Over two thirds of the subjects failed to notice the inversion (**choice blindness**: inability to recall a choice even shortly after having made it)
 - Over half of the participants **argued unequivocally for the opposite** of their original attitude in at least one instance
- Moral:

We are capable of readily providing arguments for views that we take to be our own even when we actually think they are false

The answer swap experiment



The answer swap experiment

- The experiment (Trouche *et al* 2015):
 - Subjects were asked to produce a series of arguments in response to reasoning problems, such as:

“The 4th fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?
(a) all the fruits are organic
(b) none of the fruits are organic, etc.”
 - They were then asked to assess the responses of other subjects to the same problems
 - Plot twist: among some of the responses reviewed by the subjects were *their own*, albeit presented as coming from others

The answer swap experiment

- The experiment (Trouche *et al* 2015):
 - Subjects were asked to produce a series of arguments in response to reasoning problems, such as:

“The 4th fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?

(a) all the fruits are organic

(b) none of the fruits are organic, etc.”
 - They were then asked to assess the responses of other subjects to the same problems
 - Plot twist: among some of the responses reviewed by the subjects were *their own*, albeit presented as coming from others

The answer swap experiment

- The experiment (Trouche *et al* 2015):
 - Subjects were asked to produce a series of arguments in response to reasoning problems, such as:

“The 4th fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?

(a) all the fruits are organic

(b) none of the fruits are organic, etc.”
 - They were then asked to assess the responses of other subjects to the same problems
 - Plot twist: among some of the responses reviewed by the subjects were *their own*, albeit presented as coming from others

The answer swap experiment

- The experiment (Trouche *et al* 2015):
 - Subjects were asked to produce a series of arguments in response to reasoning problems, such as:

“The 4th fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?

(a) all the fruits are organic

(b) none of the fruits are organic, etc.”
 - They were then asked to assess the responses of other subjects to the same problems
 - Plot twist: among some of the responses reviewed by the subjects were *their own*, albeit presented as coming from others

The answer swap experiment

- The experiment (Trouche *et al* 2015):
 - Subjects were asked to produce a series of arguments in response to reasoning problems, such as:

“The 4th fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?

(a) all the fruits are organic

(b) none of the fruits are organic, etc.”
 - They were then asked to assess the responses of other subjects to the same problems
 - Plot twist: among some of the responses reviewed by the subjects were *their own*, albeit presented as coming from others

The answer swap experiment

- The experiment (Trouche *et al* 2015):
 - Subjects were asked to produce a series of arguments in response to reasoning problems, such as:

“The 4th fruit and vegetable shop carries, among other products, apples. None of the apples are organic. What can you say for sure about whether fruits are organic in this shop?

(a) all the fruits are organic

(b) none of the fruits are organic, etc.”
 - They were then asked to assess the responses of other subjects to the same problems
 - Plot twist: among some of the responses reviewed by the subjects were *their own*, albeit presented as coming from others

The answer swap experiment (ctd)

- The results:
 - About half the subjects failed to notice the swap (choice blindness again)
 - Among these, more than half **rejected their own arguments** (Note: invalid arguments still more likely to be rejected than valid ones)
- Moral:

We are more critical of arguments that we take to have been provided by other people in support of their views than we are of the arguments that we provide for our own

The answer swap experiment (ctd)

- The results:
 - About half the subjects failed to notice the swap (choice blindness again)
 - Among these, more than half **rejected their own arguments** (Note: invalid arguments still more likely to be rejected than valid ones)
- Moral:

We are more critical of arguments that we take to have been provided by other people in support of their views than we are of the arguments that we provide for our own

The answer swap experiment (ctd)

- The results:
 - About half the subjects failed to notice the swap (choice blindness again)
 - Among these, more than half **rejected their own arguments** (Note: invalid arguments still more likely to be rejected than valid ones)
- Moral:

We are more critical of arguments that we take to have been provided by other people in support of their views than we are of the arguments that we provide for our own

The answer swap experiment (ctd)

- The results:
 - About half the subjects failed to notice the swap (choice blindness again)
 - Among these, more than half **rejected their own arguments** (Note: invalid arguments still more likely to be rejected than valid ones)

- Moral:

We are more critical of arguments that we take to have been provided by other people in support of their views than we are of the arguments that we provide for our own

The answer swap experiment (ctd)

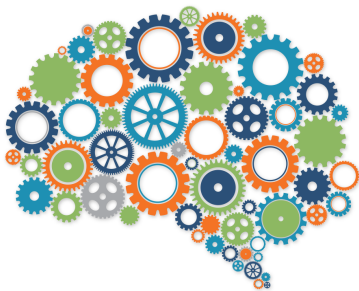
- The results:
 - About half the subjects failed to notice the swap (choice blindness again)
 - Among these, more than half **rejected their own arguments** (Note: invalid arguments still more likely to be rejected than valid ones)
- Moral:

We are more critical of arguments that we take to have been provided by other people in support of their views than we are of the arguments that we provide for our own

The answer swap experiment (ctd)

- The results:
 - About half the subjects failed to notice the swap (choice blindness again)
 - Among these, more than half **rejected their own arguments** (Note: invalid arguments still more likely to be rejected than valid ones)
- Moral:

We are more critical of arguments that we take to have been provided by other people in support of their views than we are of the arguments that we provide for our own



Critical Thinking ^ Innovation

JAKE CHANDLER

Confirmation bias & overconfidence

CONFIRMATION BIAS

Confirmation bias

- The previous experiments are thought to reveal
Confirmation bias: The tendency to search for, interpret or remember information in such a way that would tend to support, rather than undermine, an existing belief
- This is a form of **conservatism** in belief, alongside
 - Anchoring bias
 - **Primacy effects**: the tendency for evidence acquired earlier to have a greater impact than evidence acquired later

Confirmation bias

- The previous experiments are thought to reveal

Confirmation bias: The tendency to search for, interpret or remember information in such a way that would tend to support, rather than undermine, an existing belief

- This is a form of **conservatism** in belief, alongside
 - Anchoring bias
 - **Primacy effects:** the tendency for evidence acquired earlier to have a greater impact than evidence acquired later

Confirmation bias

- The previous experiments are thought to reveal
 - Confirmation bias:** The tendency to search for, interpret or remember information in such a way that would tend to support, rather than undermine, an existing belief
- This is a form of **conservatism** in belief, alongside
 - Anchoring bias
 - **Primacy effects:** the tendency for evidence acquired earlier to have a greater impact than evidence acquired later

Confirmation bias

- The previous experiments are thought to reveal
 - Confirmation bias:** The tendency to search for, interpret or remember information in such a way that would tend to support, rather than undermine, an existing belief
- This is a form of **conservatism** in belief, alongside
 - Anchoring bias
 - **Primacy effects:** the tendency for evidence acquired earlier to have a greater impact than evidence acquired later

Confirmation bias

- The previous experiments are thought to reveal
 - Confirmation bias:** The tendency to search for, interpret or remember information in such a way that would tend to support, rather than undermine, an existing belief
- This is a form of **conservatism** in belief, alongside
 - Anchoring bias
 - **Primacy effects:** the tendency for evidence acquired earlier to have a greater impact than evidence acquired later

Confirmation bias

- The previous experiments are thought to reveal
 - Confirmation bias:** The tendency to search for, interpret or remember information in such a way that would tend to support, rather than undermine, an existing belief
- This is a form of **conservatism** in belief, alongside
 - Anchoring bias
 - **Primacy effects:** the tendency for evidence acquired earlier to have a greater impact than evidence acquired later

Confirmation bias and motivation

- Confirmation bias can seemingly sometimes be motivated (i.e. due to the desire to promote one's own interests)
- Possible motivations:
 - Avoidance of losing face in public debates
 - Maintenance of beliefs that are practically useful (independently of their truth) or psychologically comforting
 - Avoidance of cognitive dissonance, discomfort resulting from holding beliefs that conflict with one's other beliefs

Confirmation bias and motivation

- Confirmation bias can seemingly sometimes be **motivated** (i.e. due to the desire to promote one's own interests)
- Possible motivations:
 - Avoidance of **losing face** in public debates
 - Maintenance of beliefs that are **practically useful** (independently of their truth) or **psychologically comforting**
 - Avoidance of **cognitive dissonance**, discomfort resulting from holding beliefs that conflict with one's other beliefs

Confirmation bias and motivation

- Confirmation bias can seemingly sometimes be **motivated** (i.e. due to the desire to promote one's own interests)
- Possible motivations:
 - Avoidance of **losing face** in public debates
 - Maintenance of beliefs that are **practically useful** (independently of their truth) or **psychologically comforting**
 - Avoidance of **cognitive dissonance**, discomfort resulting from holding beliefs that conflict with one's other beliefs

Confirmation bias and motivation

- Confirmation bias can seemingly sometimes be **motivated** (i.e. due to the desire to promote one's own interests)
- Possible motivations:
 - Avoidance of **losing face** in public debates
 - Maintenance of beliefs that are **practically useful** (independently of their truth) or **psychologically comforting**
 - Avoidance of **cognitive dissonance**, discomfort resulting from holding beliefs that conflict with one's other beliefs

Confirmation bias and motivation

- Confirmation bias can seemingly sometimes be **motivated** (i.e. due to the desire to promote one's own interests)
- Possible motivations:
 - Avoidance of **losing face** in public debates
 - Maintenance of beliefs that are **practically useful** (independently of their truth) or **psychologically comforting**
 - Avoidance of **cognitive dissonance**, discomfort resulting from holding beliefs that conflict with one's other beliefs

Confirmation bias and motivation

- Confirmation bias can seemingly sometimes be **motivated** (i.e. due to the desire to promote one's own interests)
- Possible motivations:
 - Avoidance of **losing face** in public debates
 - Maintenance of beliefs that are **practically useful** (independently of their truth) or **psychologically comforting**
 - Avoidance of **cognitive dissonance**, discomfort resulting from holding beliefs that conflict with one's other beliefs

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Confirmation bias and motivation (ctd)

- But the bias often doesn't seem to be motivated:
 - It exists for beliefs that are
 - private
 - of no practical relevance beyond their possible truth
 - psychologically uncomfortable
 - Dissonance could presumably also be avoided by changing the belief rather than ignoring the evidence

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

Why are we like this?

- In our discussion of fallacies in probabilistic reasoning, we offered a limited rationalisation in terms of the use of heuristics
- Is there a plausible story for confirmation bias?
- Mercier & Sperber (2017) claim it performs **social functions**
- They argue confirmation bias is a good way of:
 - (1) fulfilling a primary **persuasive function** of argumentation
 - (2) achieving an efficient **cognitive division of labour**

(1) Persuasive function

- *Suggestion*: Argumentation didn't evolve specifically to convince people of the *truth*, rather it evolved to simply make people share our beliefs, whether these are true or not
- *Claim*: This explains why we are not so critical of ourselves (we don't want to shoot ourselves in the foot) and much more critical of others (we want to win the debate)
- *Worry*: Wouldn't being attentive to potential objections to one's arguments make them more effective?
- *Response*: Sure, but that would take time and effort and we can simply wait for other people to provide the objections

(1) Persuasive function

- *Suggestion*: Argumentation didn't evolve specifically to convince people of the *truth*, rather it evolved to simply make people share our beliefs, whether these are true or not
- *Claim*: This explains why we are not so critical of ourselves (we don't want to shoot ourselves in the foot) and much more critical of others (we want to win the debate)
- *Worry*: Wouldn't being attentive to potential objections to one's arguments make them more effective?
- *Response*: Sure, but that would take time and effort and we can simply wait for other people to provide the objections

(1) Persuasive function

- *Suggestion*: Argumentation didn't evolve specifically to convince people of the *truth*, rather it evolved to simply make people share our beliefs, whether these are true or not
- *Claim*: This explains why we are not so critical of ourselves (we don't want to shoot ourselves in the foot) and much more critical of others (we want to win the debate)
- *Worry*: Wouldn't being attentive to potential objections to one's arguments make them more effective?
- *Response*: Sure, but that would take time and effort and we can simply wait for other people to provide the objections

(1) Persuasive function

- *Suggestion*: Argumentation didn't evolve specifically to convince people of the *truth*, rather it evolved to simply make people share our beliefs, whether these are true or not
- *Claim*: This explains why we are not so critical of ourselves (we don't want to shoot ourselves in the foot) and much more critical of others (we want to win the debate)
- *Worry*: Wouldn't being attentive to potential objections to one's arguments make them more effective?
- *Response*: Sure, but that would take time and effort and we can simply wait for other people to provide the objections

(1) Persuasive function

- *Suggestion*: Argumentation didn't evolve specifically to convince people of the *truth*, rather it evolved to simply make people share our beliefs, whether these are true or not
- *Claim*: This explains why we are not so critical of ourselves (we don't want to shoot ourselves in the foot) and much more critical of others (we want to win the debate)
- *Worry*: Wouldn't being attentive to potential objections to one's arguments make them more effective?
- *Response*: Sure, but that would take time and effort and we can simply wait for other people to provide the objections

(1) Persuasive function (ctd)

Example (from Mercier & Sperber 2017)

Consider

Hélène: We should go to Isami; it's a good restaurant.

Marjorie: I don't know. I don't have much money at the moment and Japanese restaurants can be quite pricy.

Hélène: But this one is quite cheap.

Hélène could have anticipated Marjorie's objection and dealt with it from the start but it was easier to just wait for it.

(1) Persuasive function (ctd)

Example (from Mercier & Sperber 2017)

Consider

Hélène: We should go to Isami; it's a good restaurant.

Marjorie: I don't know. I don't have much money at the moment and Japanese restaurants can be quite pricy.

Hélène: But this one is quite cheap.

Hélène could have anticipated Marjorie's objection and dealt with it from the start but it was easier to just wait for it.

(1) Persuasive function (ctd)

Example (from Mercier & Sperber 2017)

Consider

Hélène: We should go to Isami; it's a good restaurant.

Marjorie: I don't know. I don't have much money at the moment and Japanese restaurants can be quite pricy.

Hélène: But this one is quite cheap.

Hélène could have anticipated Marjorie's objection and dealt with it from the start but it was easier to just wait for it.

(1) Persuasive function (ctd)

Example (from Mercier & Sperber 2017)

Consider

Hélène: We should go to Isami; it's a good restaurant.

Marjorie: I don't know. I don't have much money at the moment and Japanese restaurants can be quite pricy.

Hélène: But this one is quite cheap.

Hélène could have anticipated Marjorie's objection and dealt with it from the start but it was easier to just wait for it.

(1) Persuasive function (ctd)

Example (from Mercier & Sperber 2017)

Consider

Hélène: We should go to Isami; it's a good restaurant.

Marjorie: I don't know. I don't have much money at the moment and Japanese restaurants can be quite pricy.

Hélène: But this one is quite cheap.

Hélène could have anticipated Marjorie's objection and dealt with it from the start but it was easier to just wait for it.

(2) Cognitive division of labour

- *Suggestion:* Confirmation bias enforces an efficient division of inquiry: different individuals “specialise” in different hypotheses

Example (from Mercier & Sperber 2017)

Imagine two engineers who have to come up with the best design for a bridge...[A]ll they want is to build a good bridge. Ella favors a suspension bridge, Dick a cantilever bridge. One way to proceed would be for each of them to exhaustively look at the pros and cons of both options, weigh them, and rate them. They would then just have to average their ratings—no discussion needed, but a lot of research. Alternatively, they can each build a case for their favored option...They would then debate which option is best, listening to and evaluating each other's arguments...[T]his option means less work for the same result: Ella and Dick each have to find only half as many arguments to thoroughly review the pros and cons of each option.

(2) Cognitive division of labour

- *Suggestion:* Confirmation bias enforces an efficient division of inquiry: different individuals “specialise” in different hypotheses

Example (from Mercier & Sperber 2017)

Imagine two engineers who have to come up with the best design for a bridge...[A]ll they want is to build a good bridge. Ella favors a suspension bridge, Dick a cantilever bridge. One way to proceed would be for each of them to exhaustively look at the pros and cons of both options, weigh them, and rate them. They would then just have to average their ratings—no discussion needed, but a lot of research. Alternatively, they can each build a case for their favored option...They would then debate which option is best, listening to and evaluating each other's arguments...[T]his option means less work for the same result: Ella and Dick each have to find only half as many arguments to thoroughly review the pros and cons of each option.

(2) Cognitive division of labour

- *Suggestion:* Confirmation bias enforces an efficient division of inquiry: different individuals “specialise” in different hypotheses

Example (from Mercier & Sperber 2017)

Imagine two engineers who have to come up with the best design for a bridge...[A]ll they want is to build a good bridge. Ella favors a suspension bridge, Dick a cantilever bridge. One way to proceed would be for each of them to exhaustively look at the pros and cons of both options, weigh them, and rate them. They would then just have to average their ratings—no discussion needed, but a lot of research. Alternatively, they can each build a case for their favored option...They would then debate which option is best, listening to and evaluating each other's arguments...[T]his option means less work for the same result: Ella and Dick each have to find only half as many arguments to thoroughly review the pros and cons of each option.

Overcoming confirmation bias

- Most effective strategy: **considering the opposite**, i.e. explicitly asking

How would I have treated this evidence **if I had held the opposite belief**? (Babcock *et al* 1997)

How would I have treated this evidence **if the evidence had gone the opposite way**? (Lord *et al* 1984)

- Both questions provably reduce confirmation bias by making one assess more fairly the evidence we take to support our beliefs

Overcoming confirmation bias

- Most effective strategy: **considering the opposite**, i.e. explicitly asking

How would I have treated this evidence **if I had held the opposite belief?** (Babcock *et al* 1997)

How would I have treated this evidence **if the evidence had gone the opposite way?** (Lord *et al* 1984)

- Both questions provably reduce confirmation bias by making one assess more fairly the evidence we take to support our beliefs

Overcoming confirmation bias

- Most effective strategy: **considering the opposite**, i.e. explicitly asking

How would I have treated this evidence **if I had held the opposite belief?** (Babcock *et al* 1997)

How would I have treated this evidence **if the evidence had gone the opposite way?** (Lord *et al* 1984)

- Both questions probably reduce confirmation bias by making one assess more fairly the evidence we take to support our beliefs

Overcoming confirmation bias

- Most effective strategy: **considering the opposite**, i.e. explicitly asking

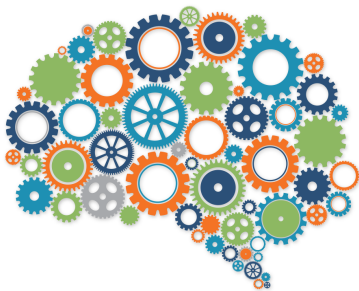
How would I have treated this evidence **if I had held the opposite belief?** (Babcock *et al* 1997)

How would I have treated this evidence **if the evidence had gone the opposite way?** (Lord *et al* 1984)

- Both questions probably reduce confirmation bias by making one assess more fairly the evidence we take to support our beliefs

Overcoming confirmation bias

- Most effective strategy: **considering the opposite**, i.e. explicitly asking
 - How would I have treated this evidence **if I had held the opposite belief?** (Babcock *et al* 1997)
 - How would I have treated this evidence **if the evidence had gone the opposite way?** (Lord *et al* 1984)
- Both questions provably reduce confirmation bias by making one assess more fairly the evidence we take to support our beliefs



Critical Thinking ^ Innovation

JAKE CHANDLER

Confirmation bias & overconfidence

CONFIRMATION BIAS AND OVERCONFIDENCE

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration

- Confirmation bias has been claimed to contribute to a further damaging and famous pathology: **overconfidence bias**
- A person is **well-calibrated** to the extent that the following match
 - their level of **confidence** (how likely they *think they are* to be right)
 - their level of **performance** (how likely *they are* to be right)

Example

If a person is well-calibrated, then we'd expect that

- 10% of all judgments assigned a 0.1 probability of being correct are correct,
- 20% of all judgments assigned a probability of 0.2 are correct, etc.

Calibration curves

- A **calibration curve** charts the frequency of correct judgments corresponding to each level of confidence
- Well-calibrated people: curve $\approx$ diagonal

Calibration curves

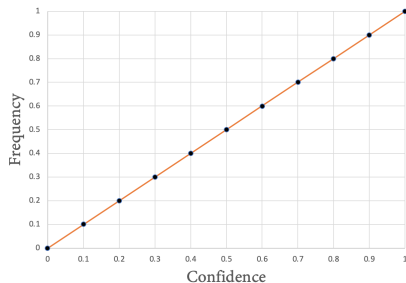
- A **calibration curve** charts the frequency of correct judgments corresponding to each level of confidence
- Well-calibrated people: curve $\approx$ diagonal

Calibration curves

- A **calibration curve** charts the frequency of correct judgments corresponding to each level of confidence
- Well-calibrated people: curve $\approx$ diagonal

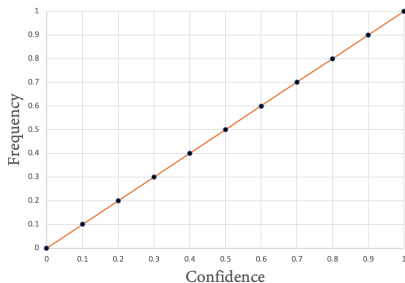
Calibration curves

- A **calibration curve** charts the frequency of correct judgments corresponding to each level of confidence
- Well-calibrated people: curve $\approx$ diagonal



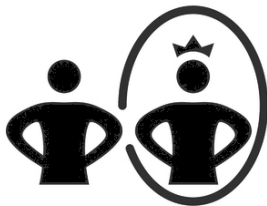
Calibration curves

- A **calibration curve** charts the frequency of correct judgments corresponding to each level of confidence
- Well-calibrated people: curve $\approx$ diagonal



- Are people well-calibrated?

Calibration experiments



Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments

- Typical experiment:
 - Subjects are presented with a series of questions
 - They are asked of each question:
 - (i) what the answer is
 - (ii) how likely they are to have gotten the answer right
- Typical results for general knowledge questions (and many others):
 - Calibration curves lie significantly below the diagonal
 - Correlation between confidence and success is often very poor
- We're often more confident than our performance warrants!

Calibration experiments (ctd)

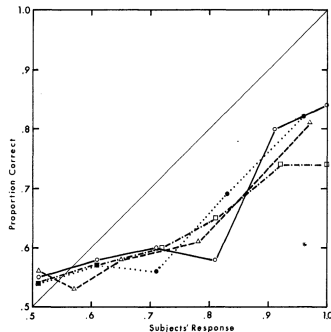
- Some curves from Lichtenstein *et al* (1975)

Calibration experiments (ctd)

- Some curves from Lichtenstein *et al* (1975)

Calibration experiments (ctd)

- Some curves from Lichtenstein *et al* (1975)



Some caveats

- Overconfidence is lower when
 - Questions are more **representative** of those that the subjects regularly encounter (Juslin *et al* 2000)
 - Subjects receive large volumes of **feedback** on their performance (e.g. meteorologists)

Some caveats

- Overconfidence is lower when
 - Questions are more **representative** of those that the subjects regularly encounter (Juslin *et al* 2000)
 - Subjects receive large volumes of **feedback** on their performance (e.g. meteorologists)

Some caveats

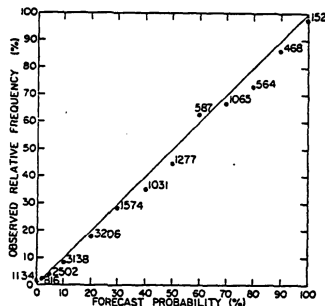
- Overconfidence is lower when
 - Questions are more **representative** of those that the subjects regularly encounter (Juslin *et al* 2000)
 - Subjects receive large volumes of **feedback** on their performance (e.g. meteorologists)

Some caveats

- Overconfidence is lower when
 - Questions are more **representative** of those that the subjects regularly encounter (Juslin *et al* 2000)
 - Subjects receive large volumes of **feedback** on their performance (e.g. meteorologists)

Some caveats

- Overconfidence is lower when
 - Questions are more **representative** of those that the subjects regularly encounter (Juslin *et al* 2000)
 - Subjects receive large volumes of **feedback** on their performance (e.g. meteorologists)



Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health:** confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image:** perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health:** confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image:** perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Why are we like this?

- Overconfidence obviously has negative repercussions
- But it could still be adaptive due to a contribution to
 - **Health**: confidence is linked to improvements in medical prospects (incl. immune response)
 - **Social image**: perceived confidence is used as a measure of competence, and affects ability to
 - influence others
 - avoid threat
 - form alliances
 - find mates

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

Overcoming overconfidence

- We can't typically get the same level of feedback as meteorologists
- What can we do to improve our situation?
- Clearly: generally **revise downwards** our initial self-confidence
- But also: explicitly **consider possible objections** to one's views
- Koriat *et al* (1980):
 - Subjects were asked questions, under different instructions:
 - Control instructions: choose an answer and estimate probability of correctness
 - Reasons instructions: as before but first list reasons for and against each of the alternatives
 - Result: Much lower levels of overconfidence under reasons instructions

References

- Babcock, L., Loewenstein, G., & Issacharoff, S. (1997). Creating convergence: Debiasing biased litigants. *Law & Social Inquiry*, 22(4), 913-925.
- Evans, J. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum
- Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE*, 7(9), e45457
- Juslin, P., Winman, A. & Olsson, H. (2000). Naive Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard–Easy Effect. *Psychological review*. 107. 384-96.

References (ctd)

- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 107.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1975) Calibration of probabilities: The state of the art. In H. Jungerman & G. de Zeeuw (Eds.), *Decision making and change in human affairs*. Amsterdam, The Netherlands: Reidel
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231-1243
- Lord, C., Ross, L., & Lepper, M. (1979). Biased Assimilation and Attitude Polarization: The effects of Prior Theories on Subsequently Considered Evidence. *JSPS*, 1979, 37, 2098-2109

References (ctd)

- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Trouche, E. *et al* (2016). The selective laziness of reasoning. *Cognitive Science*, 40, 2122–2136.