

# Common Knowledge

Puzzle 1:

Apologies

Salala Pakistan, Nov 26 2011...

# US gunships responding to request for air-support accidentally killed 24 Pakistani soldiers



Image is in the public domain.

In response...

## Pakistan closed supply routes

Image removed to copyright restrictions.

[View](#) a map of NATO supply routes through Pakistan.

Leading to ~\$1 billion in extra shipping fees



Image courtesy of [DayLove](#) on Flickr. CC BY-NC-SA

Until the US apologized...

“We are sorry for the losses suffered by the Pakistani military”

-Hillary Clinton, July 3 2012

Over half a year later!

Then, immediately...

“...the ground supply lines into Afghanistan are opening”

Why wouldn't US just say sorry?

- Mere words?

- Worth a billion dollars?

Why would Pakistan care?

- Again, mere words?

- US could just "fake it"; why "believe"?

In general, why are “symbolic gestures” such as bows, salutes, and flags so significant?

And what can we learn about symbolic gestures?

When will they be used?

What will qualify as a symbolic gesture?

Puzzle 2:

Why do we consider transgressions of **commission** worse than those of **omission**?



Notice:

Batman's intention is the same

The outcome is the same

But Batman (and presumably the viewer)  
thinks omission less bad

Matters in law too...

It is illegal for a physician to assist a patient in committing suicide

On September 17, 1998, Dr. Kevorkian administered Thomas Youk a lethal injection

Youk's family described the lethal injection as humane, not murder

But... On March 26, 1999, Kevorkian was charged with second-degree murder

In finding Kevorkian guilty, Judge Jessica Cooper wrote, *“The law prohibiting euthanasia was specifically reviewed and clarified by the Michigan Supreme Court several years ago”*

Source: [http://en.wikipedia.org/wiki/Jack\\_Kevorkian#Trials](http://en.wikipedia.org/wiki/Jack_Kevorkian#Trials)

But it isn't illegal for a physician to turn off life support...

Supreme Court of New Jersey.

### In the Matter of Karen QUINLAN, An Alleged Incompetent.

Argued Jan. 26, 1976. | Decided March 31, 1976.

Father sought to be appointed guardian of person and property of his 21-year-old daughter who was in a persistent vegetative state and sought the express power of authorizing the discontinuance of all extraordinary procedures for sustaining daughter's vital processes. The Superior Court, Chancery Division, 137 N.J.Super. 227, 348 A.2d 801, denied authorization for termination of the life-supporting apparatus and withheld letters of guardianship over the person of the incompetent, and father appealed and the Attorney General cross-appealed. The Supreme Court, Hughes, C.J., held that a decision by daughter to permit a noncognitive, vegetative existence to terminate by natural forces was a valuable incident of her right to privacy which could be asserted on her behalf by her guardian; that the state of the pertinent medical standards and practices which guided the attending physicians who held opinion that removal from the respirator would not conform to medical practices, standards and traditions was not such as would justify court in deeming itself bound or controlled thereby in responding to case for declaratory relief; and that upon the concurrence of guardian and family, should the attending physicians conclude there was no reasonable possibility of daughter's

physicians should consult with hospital ethics committee and if committee should agree with physicians' prognosis, the life-support systems may be withdrawn and said action shall be without any civil or criminal liability therefor, on the part of any participant, whether guardian, physician, hospital or others.

© Supreme Court of New Jersey. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

And people's moral intuitions treat omission differently from commission...

“Subjects read scenarios concerning pairs of options. One option was an omission, the other, a commission. Intentions, motives, and consequences were held constant. Subjects either judged the morality of actors by their choices or rated the goodness of decision options. Subjects often rated harmful omissions as less immoral, or less bad as decisions, than harmful commissions. “

-Spranca, Mark, Elisa Minsk, and Jonathan Bacon. “[Omission and Commission in Judgment and Choice](#).” *Journal of Experimental Social Psychology* 27, no. 1. (January 1991): 76-105.

Even uber-rational economists treat omission and commission differently...

“One of the strictest rules in experimental economics is that the researcher may not deceive their participants.”

Croson “The Method of Experimental Economics”

# Why?

“... the validity of an economic experiment rests on the link between behavior and payoffs (incentives). If that link is weakened, the experiment becomes an inferior test of the economic theory it is designed to address.”

“... will begin to distrust the experimenter’s statements. This lack of trust could lead the participants to change their behavior in future experiments.”

# But comes at a cost

“... deception about the purpose of the experiment can aid in honest elicitation and overcome presentation effects“

“... deception is often used to examine situations which would not occur naturally, for example, how individuals respond to low ultimatum offers.”

How do economists get around the rule against lies of commission?

“That said, many of these benefits arising from deception can be enjoyed by simple omission (not informing the participants of the subject of the experiment, or doing so only very generally) rather than by commission (explicitly lying to the participants). “

Does the lie of omission have any differential effect on subject trust or subject incentives?

So why do economists differentiate between omission and commission?

In general, why do we treat omission differently from commission?

What do we learn from this? Any policy implications?

Puzzle 2:

“Categorical” Norms

Recall the norm against chemical weapons...  
See [The Oatmeal comic](#) about chemical weapons.

Seems weird that Obama, the French, and the British were poised to attack Assad only after he used chemical weapons, when he had wantonly killed many more civilians beforehand

Why didn't Obama threaten to attack after, say, 100,000 civilian casualties?

We aren't the only ones who noticed this...

“Why, it is fair to ask, does the killing of 100,000 or more with conventional weapons elicit little more than a concerned shrug, while the killing of a relative few from poison gas is enough to trigger an intervention?”

-NY Times, Steven Erlanger, Sept 6 2013

"There have been an estimated 100,000 people killed in Syria's civil war, almost all by conventional bombs and weapons. If that isn't enough to mandate a U.S. intervention, then why is the alleged use of chemical weapons, killing a mere fraction of the total, enough to make intervention obligatory?"

-Washington Times, John Glaser, Aug 22 2013

"As critics have argued for a century or more, the taboo is not rational. Chemical weapons are insidious and ghastly, yes, but so are all sorts of other ways of killing and wounding—and many of those other ways are a lot harder to defend people against,"

-The Economist

"One constituent said to me: 'It is horrendous that these children were killed, but they are being killed in other ways also. What's the difference?'"

Senator Susan Collins, Republican of Maine,  
quoted in NY Times, Aug 31 2013

So why “no chemical weapons” and not “no killing 100,000 civilians”?

How 'bout these possible explanations...

Maybe chemical weapons are a particularly painful way to die?

Maybe chemical weapons stand-in for “wanton killing of civilians”?

To rule out these alternative explanations, we turn to another time when this norm showed up...

“A strong military case was made for the use of gas before America’s attack on the island of Iwo Jima; Japanese defenders in caves and tunnels would have been particularly vulnerable. Franklin Roosevelt rejected the idea.”

-The Economist, The History of Chemical Weapons

Instead, "... flamethrowers were used to kill Japanese holed into pillboxes, buildings and caves. "

-Wiki entry for Battle of Iwo Jima



Image is in the [public domain](#).

This had a high cost...



Image is in the [public domain](#).

>26,000 American casualties, Iwo Jima, WWII

So did we avoid chemical weapons because they were more humane?

Video Clip: Japanese Soldier Burnt Alive by Flame Thrower

Uh... we don't think so

Was it because of concerns for civilians?

“Because all the civilians had been evacuated, there were no civilian casualties at Iwo Jima”

Nope.

So “no chemical weapons” isn’t because:

Chemical weapons are a particularly painful way to die

Chemical weapons stand-in for “wanton killing of civilians”

Then what is it?

In general, why do we rely on “categorical” norms?

Why is it “murder” regardless of whether victim was 2 or 82?  
Why is Obama African American, and not half African American?

Can our game theory explanation teach us something new?

When will we rely on categorical norms?  
When should we try to overcome them?  
How can we enforce more efficient norms?

Puzzle 4:

Why do we speak **indirectly**?



Why use innuendo if the cop clearly knows that a bribe was offered?

To rule out some alternative explanations (like perhaps the briber wanted to make his bribe not so clear, so that only bad cops would notice)

Let's consider another example of indirect speech...

On July 22, 1968, as El-Al flight 426 to Rome was approaching the coast of Italy, two terrorists burst into the cockpit and directed the captain to redirect the plane to Algiers.



Image courtesy of [AF1621](#). CC BY.

El-Al Boeing 707

A day after the plane landed in Algiers, the international passengers were released

But the Israeli passengers and the crew were held hostage.

Soon, negotiations started for their release...

“The diplomatic relations between Italy and Algeria had been good, unlike those between Algeria and France, Britain, and the USA. The Italians considered Algeria a “good neighbor,” so, immediately following the hijacking, the Italians conveyed a complaint via diplomatic channels. Then, when Israel approached Italy (the Israeli Ambassador was on vacation and therefore the Israeli Foreign Ministry sent Aviezer Chelouch, an old acquaintance of mine, to Rome), the Italians immediately agreed to mediate between Israel and Algeria. I was of course ignorant of all this and only told of it after my release.”

- Oded Abarbanell, Captain

Then, 40 days later...

“At seven o'clock next morning I awoke from a deep sleep: It was Lashmi who arrived and woke all of us up... and asked us to pack our belongings and get dressed - **we were being released**... 5 minutes later, all of us were ready in the corridor. Lashmi returned and told us to load our bags on the minibus. We drove to our usual meeting place and met the Italian Ambassador, the First-Secretary of the Italian Embassy and the Red-Cross representative. The usual Algerian representatives were there, as well as a crew of the Algerian television and radio. The Italian Ambassador asked if we were all well, and told us that in an hour we would depart with the regularly scheduled "Alitalia" flight to Rome. **He told me that the Italians were the mediators between Israel and the Algerians and that they conducted the entire negotiations.** I gathered from the very short conversation which I had with him that he had maintained a daily interest in our welfare and was updated constantly by the Algerian authorities, so had no need for direct contact.”

Why go through the Italians?

It's not like Israel doesn't know Algeria is offering to negotiate. And vice versa

Can we use game theory to learn...

What kinds of speech is indirect?

When is indirect speech useful?

Today...

We'll give an informal answer for apologies

If time, we'll informally show how this answer applies to other puzzles

In future classes...

We'll formalize our answers to the puzzles

Generalize to more applications

Provide evidence where available

Discuss prescriptions

Let's start with the answer to apologies...

Quick reminder of the puzzle...

# US gunships responding to request for air-support accidentally killed 24 Pakistani soldiers



Image is in the public domain.

## Pakistan closed supply routes

Image removed to copyright restrictions.

[View](#) a map of NATO supply routes through Pakistan.

Until the US apologized...

“We are sorry for the losses suffered by the Pakistani military”

-Hillary Clinton, July 3 2012

Why did these “mere words” matter?

Our explanation (has three parts)...

First...

U.S.-Pakistan relationship can be modeled as  
Hawk-Dove

Second...

Recall, we sometimes condition our behavior in Hawk-Dove

Who arrived first?

Who found the object?

This was called an “uncorrelated asymmetry”

It didn't effect payoffs

Or provide better information

But it did have the following property...

It created “common knowledge”



And...

We can condition our behavior in games like Hawk-Dove only on events that create “common knowledge”

This is our key CK theorem (which we will use over and over and over...and we will formalize in future lecture)

(Notice this answers our question from an earlier class: what can act as an uncorrelated asymmetry)

(Of course, as usual, people don't consciously think about what others are thinking they are thinking others are thinking they are thinking...

More on this in a future class)

Why do we need common knowledge to condition our behavior in games such as Hawk-Dove?

When both condition behavior on some event, E

E.g.,

I play H when I believe E occurred, D o/w

You play D when you believe E occurred, H o/w

If no CK

E.g., I believe E occurred, but think you don't believe E occurred

Then I think you will play H, so I play D

don't Or... I believe E occurred, I think you believe E occurred, but I think that you think I believe E occurred

Then I think you think I will play D, so I think you will play H, so I play D

Etc.

Third...

An apology creates common knowledge

That's because an apology has the following property

Property: Whenever I think an apology was made, I think you think an apology was made

Here's how this property leads to common knowledge...

If apology has the property:

Whenever I think an apology was made, I think you think an apology was made

By symmetry, when you think an apology was made, you think I think an apology was made

Let's put that together...

I think that you think that I think that an apology was made

Do that logic again as many times as you want...

I think that you think that I think that you think that I think that you think that I think that you think... that an apology was made.

I.e., common knowledge that an apology was made

An apology is an example of an *evident event*

Evident events have the property:

Whenever I think the event occurred, I  
think you think the event occurred

In general: Evident events → common knowledge

## Other examples of evident events...

Handshakes

Eye contact

Etc.

Whenever I think we shook hands, I think you  
think we shook hands

Whenever I think we made eye contact, I think  
you think we made eye contact

Now, our explanation, all on one slide...

- 1) U.S.-Pakistan relationship can be modeled as Hawk-Dove
- 2) We can condition our behavior in games like Hawk-Dove only on events that create “common knowledge”
- 3) Apology is an evident event, and thus creates common knowledge

Let's take a moment to consider some additional  
“symbolic gestures”

Then we'll see if the same argument applies to  
them

For example, why do we care about bowing?

# In the 10 commandments...

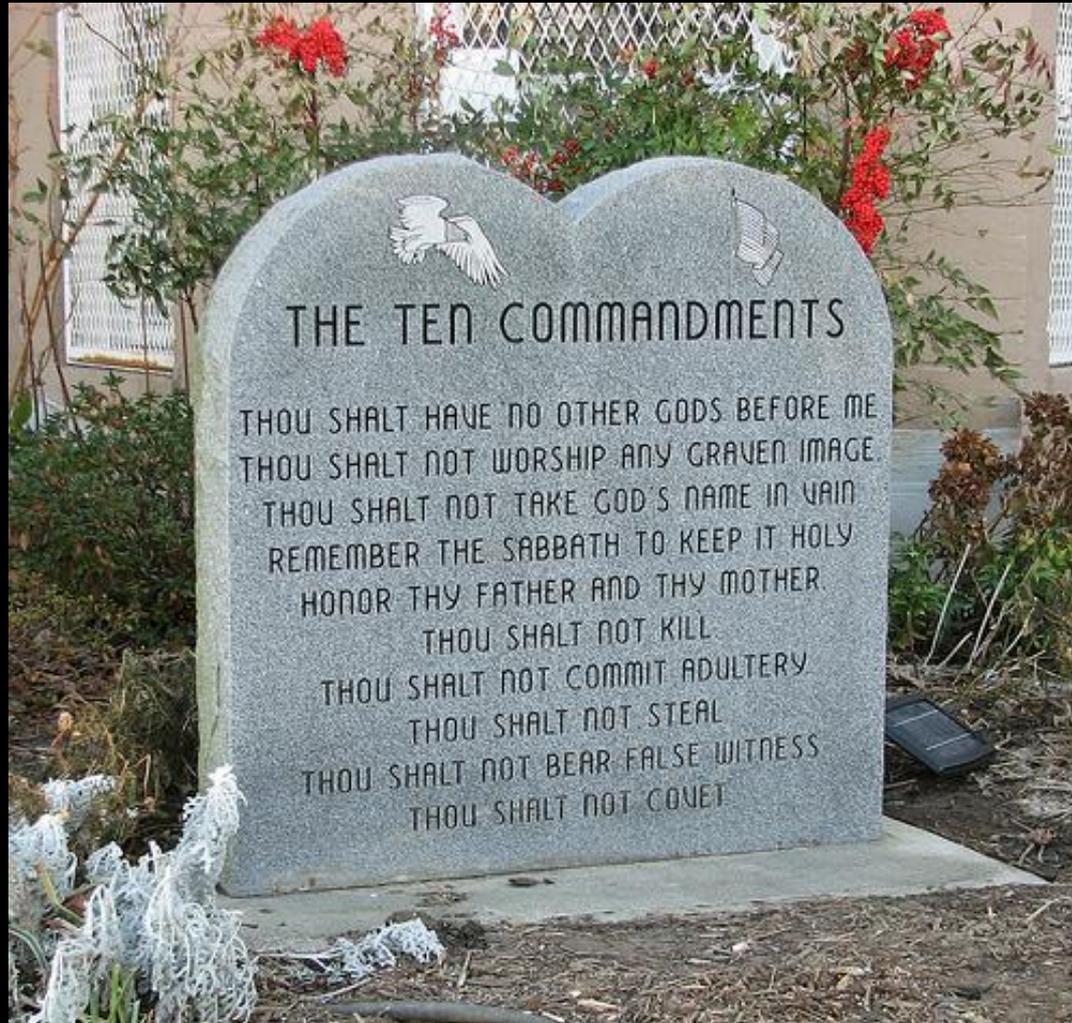


Image courtesy of [Beechwood Photography](#) on Flickr. CC BY-NC-SA

“Do not have any other gods before Me. *Do not represent [such] gods by any carved statue or picture of anything in the heaven above, on the earth below, or in the water below the land.* **Do not bow down to [such gods] or worship them.** I am God your Lord, a God who demands exclusive worship. Where My enemies are concerned, I keep in mind the sin of the fathers for [their] descendants, to the third and fourth [generation]. But for those who love Me and keep My commandments, I show love for thousands [of generations].”

(Exodus 20:3-6)

God's not the only one who cares about  
bowing...

The Romans (purportedly) killed Chana and her seven sons when Chana refused to bow to an idol ([Tractate Gittin 57b](#))



Why did God care so much?

Why would the Romans care?

Some more examples of symbolic gestures that people really care about...



Image is in the [public domain](#).



Image courtesy of [CherryPoint](#) on Flickr. CC BY-NC-SA

“Michelle Obama Hugs Queen Elizabeth, Stirs Controversy.”

-Huffington Post, May 3, 2009.



Another example... The Warsaw Uprising



Image is in the [public domain](#).

The rebels placed two flags high above the square...



The Reichsfuehrer [Himmler] bellowed into the phone: ‘Stroop, You must at all costs bring down those two flags!’”

—Jürgen Stroop, 1949

Source: Who Defended The Warsaw Ghetto? (*The Jerusalem Post*), Moshe Arens

Why would the Nazis care so much about flags?

Some more examples of symbolic gestures...

First step in Israel-Egypt  
peace process was  
Anwar Sadat's visit to  
the Knesset

Why did Israel care?



Image is in the [public domain](#).

More generally...

The main features of the treaty were **mutual recognition**, cessation of the state of war that had existed since the 1948 Arab–Israeli War, normalization of relations and the **complete withdrawal by Israel of its armed forces and civilians from the Sinai Peninsula** which Israel had captured during the Six-Day War in 1967.

-Source: [http://en.wikipedia.org/wiki/Israel-Egypt\\_Peace\\_Treaty](http://en.wikipedia.org/wiki/Israel-Egypt_Peace_Treaty)

Why does Israel care so much about  
“recognition”?

Some more examples...

“I love you”

“Please”

“Thank you”

“Sir”

What about this one...

# The tree that nearly started a war



A 100-foot poplar tree blocked the line of sight between UN command checkpoint 3 and observation post 5.

Two UNC soldiers sent to trim the tree.

But this tree had “symbolic meaning”: Kim Sung planted it

So the N. Koreans couldn't just let it be cut down...So they almost started a war to protect this “symbolic tree.”

North Korean soldiers ordered UNC to cease

UNC ordered to continue

North Korean forces killed two UNC soldiers  
(U.S.!!)

That's already a lot of death for a symbolic tree...

But now the U.S. has to send a symbolic gesture that they can't be messed with!

So the US launched “Operation Paul Bunyan”

(We’re not kidding)

US sent two eight-man teams of military engineers (from the 2nd Engineer Battalion, 2nd Infantry Division) equipped with chain-saws to cut down the tree.

With...

64-man South Korean special forces  
20 utility helicopters and 7 Cobra attack helicopters

Oh, and...

Nuclear-capable B-52 Stratofortresses  
Escorted by U.S. F-4 Phantom and South Korean F-5 Freedom Fighters and F-111 bombers

And... The aircraft carrier *Midway*

And...

Second Battalion, 71st Air Defense Regiment armed with HAWK missiles, and armor  
12,000 additional troops were ordered to Korea

Funny name for an operation, but certainly NOT A JOKE!

All to remind the North Koreans that we can trim any tree we darn well please!

Last example...

“Yasukuni shrine visits: Japan honoring the dead or insulting the neighbors?”

-CNN, October 21, 2013.

OK, here's how our explanation applies to all these gestures

First...

Recall that in apologies case, we modeled game as Hawk-Dove

Similar in all these cases, except might not be H-D.  
(need only be SOME game with multiple equilibria)

For example...

When flying a flag in Warsaw, coordinating on whether to rebel

What about the others?

Can you figure out how to model them as a game with multiple equilibria?

Next part of argument...

Like apology, other symbolic gestures are evident events

If I think someone bowed/saluted, then I think you think she bow/salute

When I see a flag, I think it's likely you saw it

Same with the others...

When we say, “Thank you” or “I love you,” we think others think we said it

When I see Arafat walk through the door first, I think you saw Arafat walk through the door first

When we think someone “recognizes” a country, chops down a tree, or visits a shrine, we think others think it happened.

The entire argument on one slide...

- 1) Each situation can be modeled as game with multiple equilibria
- 2) Can only condition behavior on event if event creates common knowledge
- 3) Symbolic gesture is an evident event
- 4) Evident event creates common knowledge

Is all this emphasis on symbolic nonsense?  
Or does it actually matter?

**IT MATTERS!** Because will condition future  
behavior on it

When do symbolic gestures matter?

Whenever situation can be modeled as game  
with multiple equilibria

What events can be “symbolic”  
(i.e. can matter even if don’t directly influence payoffs)

Any evident event!

Meaning, symbolic will have the following features:

public, physical contact, eye contact, explicit...

As compared to

rumors

innuendos

killing more than 100,000 people

omission

Which are NOT evident!

(this, of course, will help us explain those puzzles!)

A rumor

A rumor has the feature that I think someone said it, but I am not sure if you think someone said it

Or am not sure you think I think someone said it

Or...

## A lie of omission

I Even if I know intention and consequence was to lead other to act upon a false belief, for my betterment, may doubt that others realize this.  
(we will provide evidence for this!)  
Suppose punishment needs to be coordinated  
(think about how we justified punishment in our repeated games setup)  
(we will justify this with experiments)  
Then cannot start punishing after act of omission

## A norm against killing 100,000 civilians

Even if I am sure Assad killed more than 100,000 civilians, I may not be confident you are sure, especially if I estimate he killed 101,000!  
(trick: that's enough! You will see!)

## An innuendo

Same as omission... I think bad thing said, but am not sure if you think bad thing said  
We will argue, innuendo allows information to be conveyed...  
While also not allowing me to switch equilibria in a game with multiple equilibria!  
(have your cake and eat it too!)

Let's discuss in more depth omission-commission...

Recall the argument...

We punish based on “moral intuitions”

Suppose punishment needs to be coordinated  
(Think about how we justified punishment in our  
repeated games setup)

Then cannot start punishing after act of omission since omission does not create common knowledge

(So our moral intuitions will lead us to treat acts of omission less severely than acts of commission)

Now we're ready to see some evidence

## Aside

I will go through experimental design and results SLOWLY! )

- lecture on evidence that omission commission explained by higher order beliefs...
- ALSO lecture on DESIGN test game theory and social behavior!
- and GENERAL experimental design/analysis
  
- stop me if term unclear.
- Or if motivation unclear
- Or if interpretation unclear

Two studies we're working on right now

This is with Jillian Jordan, an awesome grad student at Yale

Feedback ~~welcome~~ ~~encouraged~~ needed!

First study will show that our moral intuitions are affected by what we think other potential punishers think

Second study will show that with omission we think other potential punishers are less likely to know transgression occurred, even when we are equally as sure

And that this “second order belief” explains a large part of the omission commission distinction

Study 1's design...

## Mturk study

We describe variations of the following situation to hundreds of individuals who take a survey for \$.50 on mturk.com

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

**You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.**

**Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.**

**The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.**

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

**You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.**

**Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.**

**The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.**

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two colleagues, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

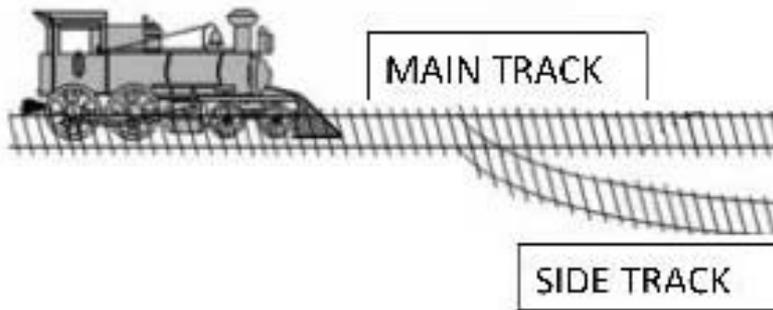
There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.

Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.

The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.



Control Buttons



You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.

Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.

**The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.**

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

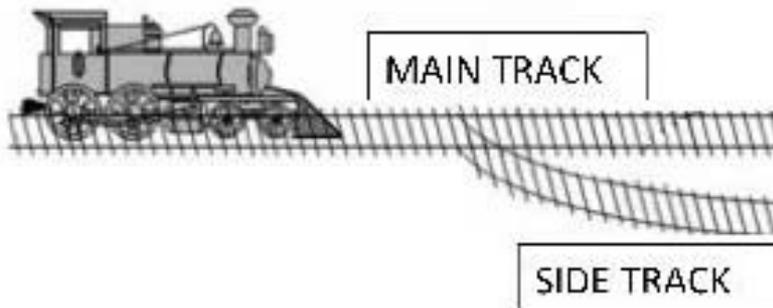
There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.

Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.

The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.



You, David, and Mark all CAN see the tracks through the window

Control Buttons



Mark



You

Mark CAN see David and the Control Buttons through the Glass Wall

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

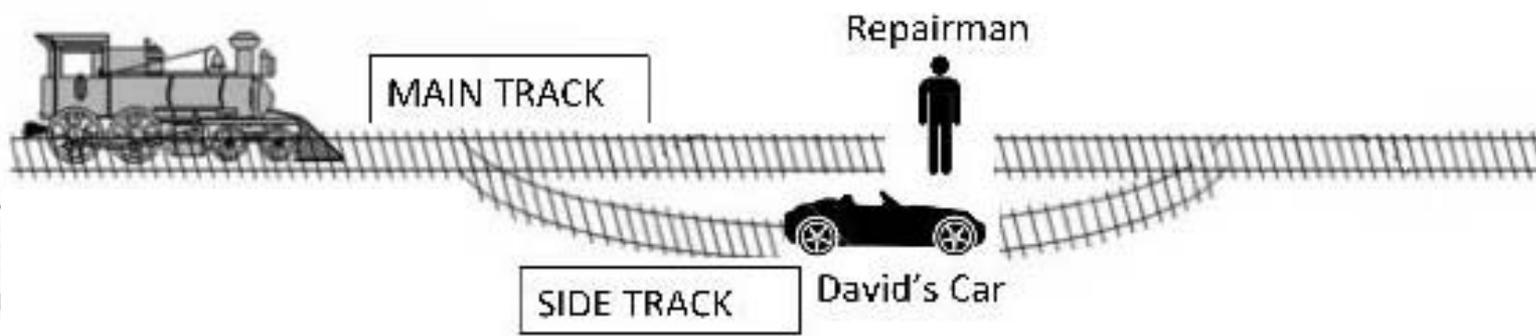
There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

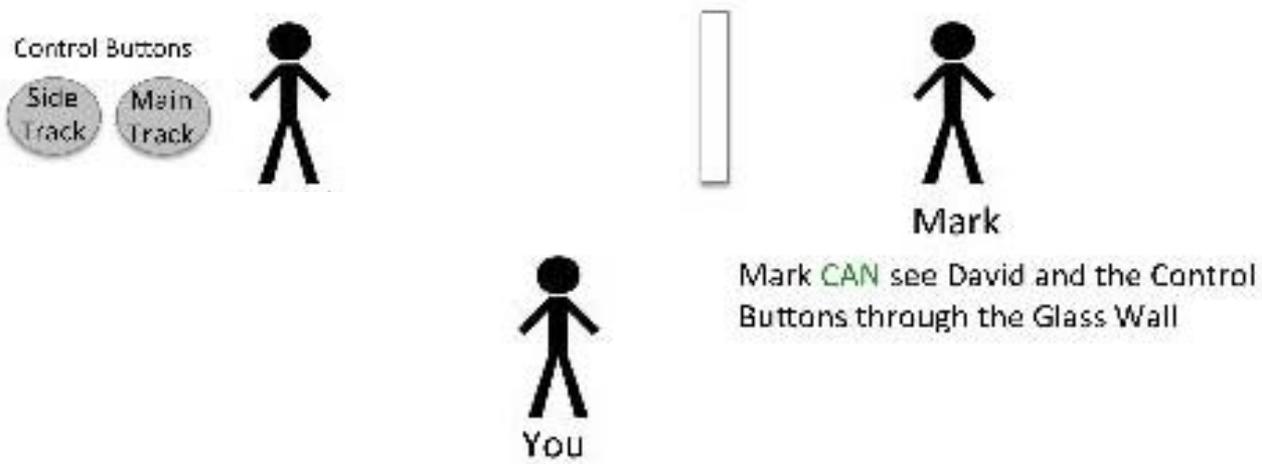
You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and examines them.

Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.

The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.



You, David, and Mark all **CAN** see the tracks through the window



You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and heading towards the main track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

**You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.**

**Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.**

**The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.**

To summarize...

You and Mark both observe David in an act of omission

Next, we ask two questions intended to elicit moral judgments...

How morally wrong do you think David's behavior is?

Please answer this question on a 1 to 100 scale, based on these **sample ratings:**

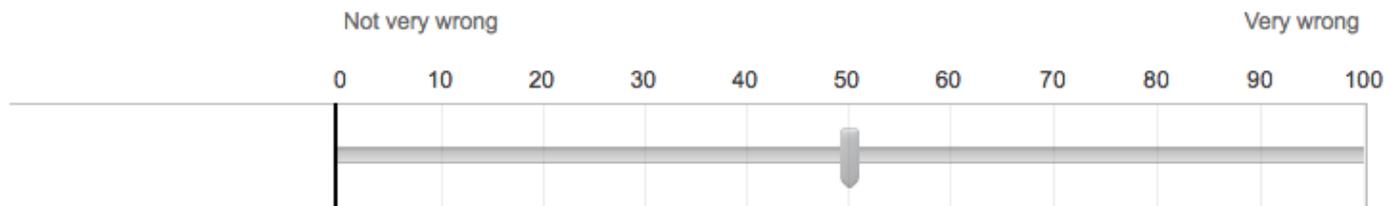
**White lie (to avoid hurting someone's feelings): 1**

**Lying for personal gain: 10**

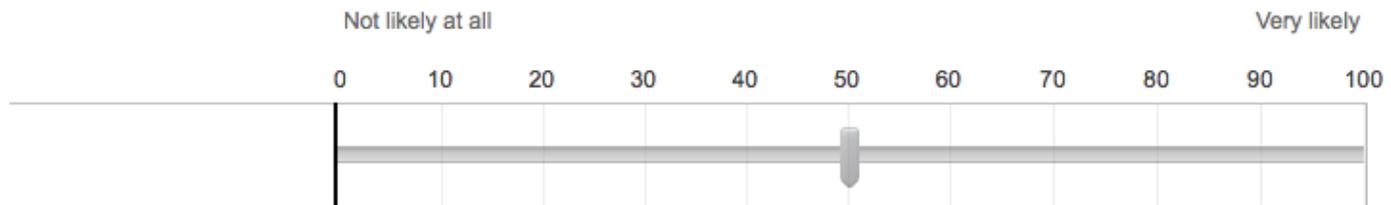
**Stealing: 30**

**Assault: 50**

**Murder: 70**

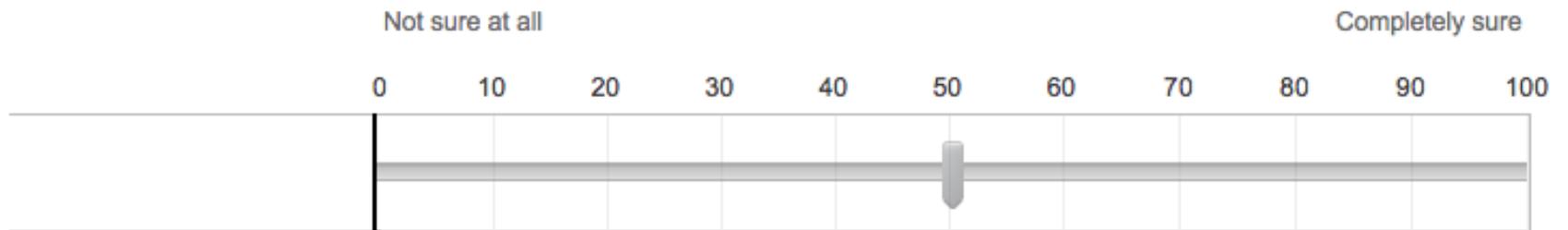


Imagine that you and David are friends. How likely would you be to ostracize David?

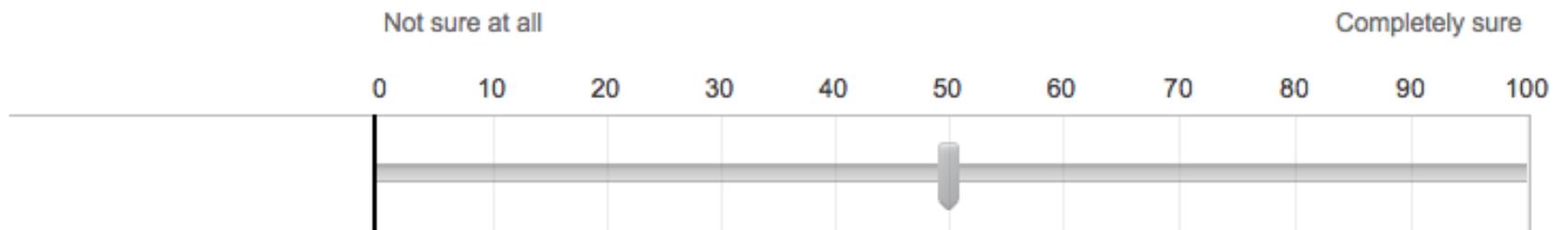


Questions designed to elicit first order beliefs...

How sure are you that David expected the repair man to be injured?



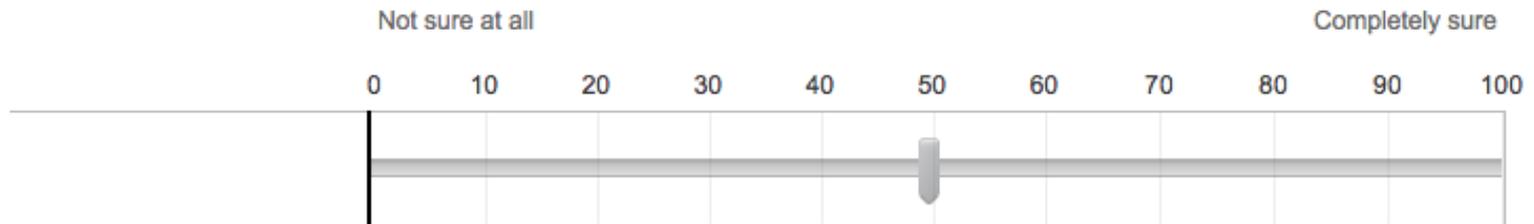
How sure are you that David preferred to save his car than to prevent the repair man's injury?



Questions designed to elicit your second order beliefs  
(your beliefs of Mark's beliefs)

Now, we would like to know what you think is going through Mark's head. When providing your answers, please take your best guess as to what Mark will find out about what happened.

How sure is Mark that David expected the repair man to be injured?



How sure is Mark that David preferred to save his car than to prevent the repair man's injury?



And questions about Mark's moral judgments...

How morally wrong does Mark think David's behavior is?

Please answer this question on a 1 to 100 scale, based on these **sample ratings:**

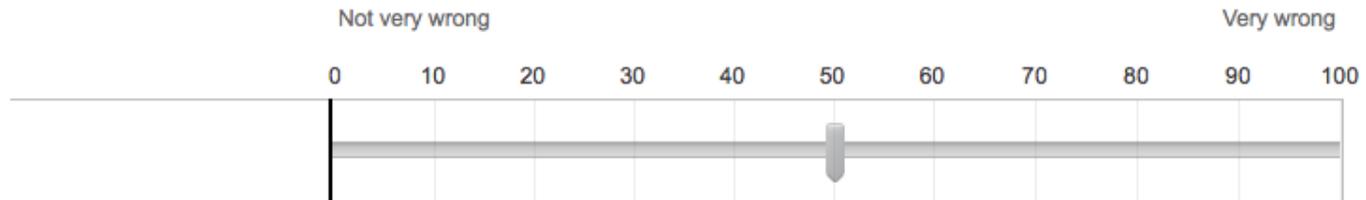
**White lie (to avoid hurting someone's feelings): 1**

**Lying for personal gain: 10**

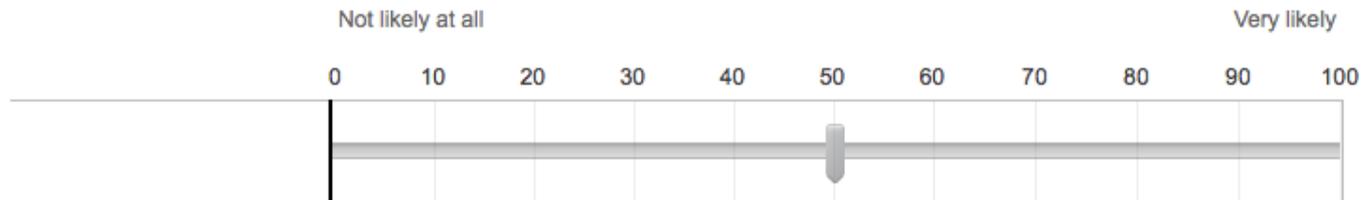
**Stealing: 30**

**Assault: 50**

**Murder: 70**



Imagine that Mark and David are friends. How likely would Mark be to ostracize David?



Now, our manipulation...

We vary whether Mark can see David...

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **However, a black wall is separating Mark from David, so he CANNOT see David and the buttons.**

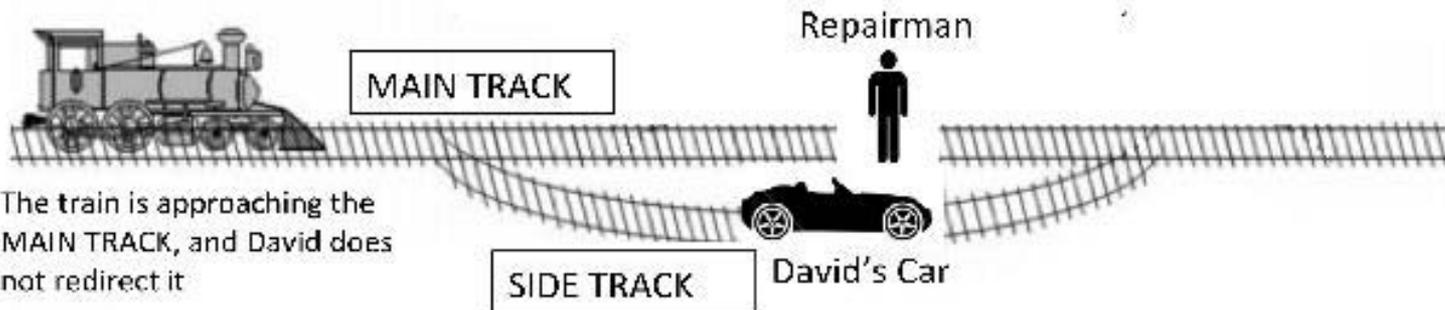
There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **However, the camera is switched OFF today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Main Track. If the train continues onto the Main Track, David's car will be fine, but the repair man will be injured. In contrast, if the train switches to the Side Track, the repair man will be fine, but David's car will be destroyed.**

**You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.**

**Finally, he sits down, never having pressed either button. As you watch David, you look over and notice that Mark CANNOT see what you are seeing, because the wall is blocking his view of David.**

**The train continues away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.**



The train is approaching the MAIN TRACK, and David does not redirect it



You, David, and Mark all **CAN** see the tracks through the window

Control Buttons



David



Mark

David does **not** press any button



You

Mark **CANNOT** see David and the Control Buttons through the Black Wall

Then we ask the same questions about moral judgments and beliefs

All subjects shown both scenarios

Mark observes

Mark doesn't observe

We compare each subject's judgments across treatments  
(within subject)

We also double check results by comparing first treatment  
across subjects  
(between subject)

What do we find?

First let's confirm that our manipulation had the desired effect on second order beliefs

This is called a “manipulation check”

(If this fails, our study does not disprove our theory, it just doesn't test our theory)

Compared to the “Mark can’t see David” condition, subjects in the “Mark can see David” condition

the thought that Mark was more sure that David expected the repairman to be injured ( $p < .001$ ) and that Mark was more sure that David preferred to save his car than to save the repairman ( $p < .001$ )

Thus, our manipulation successfully influenced subjects’ assessment of other potential punisher’s beliefs

Now we're ready to look at main results. I.e., how did second order beliefs change subjects' moral judgments?

Let's look within subject first...

As predicted:

Subjects judged David's behavior to be more wrong in the “Mark can see David” condition than the “Mark can’t see David” condition ( $p=.034$ )

This demonstrates that perceived moral wrongness is influenced by beliefs of other potential punisher’s beliefs

I.e. higher order beliefs influence moral judgments!  
(What other than our theory can explain this result?)

We found additional evidence to support the role of second order beliefs on moral judgments

Within each condition

subjects who thought Mark would find David's behavior more wrong also found David's behavior more wrong ( $p < .001$ )

When controlling for their own beliefs about David

More evidence that moral judgments are influenced by higher order beliefs!

Let's move on to study 2

First, let's recall the motivation...

We want to show

Actions are judged worse than omissions

BECAUSE

even holding constant first order beliefs  
about intention,

omission induces lower higher order beliefs

How should we do this?

Let's redo study 1, but...

Mark is NEVER present

We manipulate whether subjects are shown an act of commission vs. act of omission

How do we run the “commission” condition?

You work at a train station, illustrated below.

At this station, two buttons control a train that can travel along a Main Track or a Side Track. The "Main Track" button routes the train to the Main Track, and the "Side Track" button routes the train to the Side Track.

You work with two co-workers, David and Mark. All three of you are in a building behind the tracks. You can all see the tracks through a glass window.

David is standing next to the control buttons. You and Mark are out of reach of the buttons. You can see both Mark and David. **Mark CAN also see you, David, and the buttons through a clear glass wall.**

There is also a surveillance camera that normally films activity on the tracks and near the control buttons. **The camera is switched ON today.**

You look at the tracks and notice that neither track is clear. On the Main Track, a repairman's leg is caught. On the Side Track, David's new car is parked. **A train is approaching and routed towards the Side Track. If the train continues onto the Side Track, the repair man will be fine, but David's car will be destroyed. In contrast, if the train switches to the Main Track, David's car will be fine, but the repair man will be injured.**

You see David looking conflicted. He glances back and forth between the two buttons, and the two tracks. He looks at the repair man for a long time, and at his new car for a long time. He then takes his car keys out of his pocket and admires them.

Finally, he presses the "Main Track" button. As you watch David, you look over and notice that Mark is watching David too. You and Mark make eye contact as you watch David make his decision.

The train redirects away from David's car and towards the repair man, hitting him. The repair man lives, but his leg is badly broken.

What happened?

(Here we just analyze data within-subject, due to smaller sample sizes)

We reproduce the standard effect that subjects judge actions to be more wrong than omissions ( $p < .001$ )

We ALSO find...  
(this part is cool!)

This difference persisted after “controlling for” subjects' own sureness/wrongness ratings

Thereby ruling out alternative explanation that omission-commission effect driven SOLELY by “differences in first order beliefs”  
(i.e. treat commission different because more sure about intentions)

And...

(here's the FIRST novel result!)

Subjects thought that **Mark** would find David's behavior more wrong in the action than omission condition ( $p < .001$ )

Even when controlling for subjects' own sureness/wrongness ratings

Thus, actions and omissions appear to differ in higher order beliefs, even when controlling for first order beliefs

Likewise...

Compared to the omission condition, subjects in the commission condition thought Mark would be more sure that David expected the repairman to be injured ( $p < .001$ )

So we showed...

Our moral intuitions are affected by what we think other potential punishers think  
(study 1)

With omission we think others are less likely to know transgression occurred, even when we are equally as sure  
(what we just used study 2 to show)

Last thing:

Let's show that the difference in second order beliefs EXPLAINS the omission commission distinction!

(well, explains PART OF...)

We show this using a statistical trick...

When predicting the effect of commission on moral wrongness,

If we add controls for third party belief and “wrongness attributions”, the “coefficient” on the “commission dummy” drops from 12.26 ( $p < .001$ ) to 7.38 ( $p < .001$ )

Further studies?

Perhaps incentivized lab experiment instead of mturk surveys?

Perhaps realistic scenarios instead of contrived “trolley problems”?

Perhaps manipulate evident (e.g. manipulate whether make eye contact with Mark instead of whether Mark can see?)

Perhaps directly manipulate report of second order beliefs (e.g. Mark sees same thing but in one case Mark judges intentions to be worse)

Notice:

In signaling games higher order beliefs don't matter.

All that matters is player 2's beliefs about 1's type.

Doesn't matter what 2 thinks 1 thinks or what 1 thinks 2 thinks etc.

→ One more prediction from our model that hard to reconcile with alternative explanations (and hence would be good evidence):

omission/commission will matter when judgments of others driven by “repeated games”, not “signaling games”

(perhaps when deciding who to punish vs. deciding who to interact with?)

(How test? Same as we had, just change the d.v.!)

Any other novel predictions?

Let's switch gears and discuss policy implications...

Why are judges/lawyers/policy makers affected by omission/commission?

Claim:

spillover from “moral intuitions” learned/evolved where important to coordinate punishment.

But NOT important for law to coordinate!

So policy makers/lawyers/judges SHOULD ignore!

(i.e. we suggest removing legal distinction between physician assisted suicide and cessation of treatment!)

(i.e. we suggest economics journals should NOT treat lies of commission different from lies of omission!)

(i.e. Kant’s differential treatment of omission vs. commission has more to do with how his moral intuitions are learned/evolved than “objective truth”

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.11 Insights from Game Theory into Social Behavior  
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.