

“IT’S NOT ABOUT THE MONEY.
IT’S ABOUT SENDING A MESSAGE!”

BELIEF-BASED MOTIVES BEHIND PUNISHMENT

Andras Molnar, Shereen Chaudhry, & George Loewenstein



MOTIVATION

Why do we care whether “bad guys” know the reason why they are punished, even if they have already gotten their just desserts?

Why is it so dissatisfying to see them suffer without understanding, even if we know that they deserve their punishment?

Are punishment decisions driven by abstract principles: such as considerations for distributive and retributive justice, or is there something else that punishers also care about?

In other words, do punishers care about what transgressors **believe**, in addition to what they get and how they feel?



*A climactic moment of film history:
the villain, Frank, just realizes who, and for
what reason, is taking revenge on him.
(Once Upon a Time in the West, 1968)*

WHY DO PEOPLE PUNISH?

The literature on punishment has focused mainly on two questions:

1) Whether punishment is:

- the means to achieve some other goal, or
- an end in itself.

2) Whether punishment is:

- “cold” – strategic, deliberate, and calculated, or
- “hot” – an almost automatic affective response, driven by anger.

Carlsmith et al., (2002)
Fehr & Gächter (2002)
Darley & Pittman (2003)
Henrich et al. (2006)
Carlsmith et al. (2008)
Strelan & van Prooijen (2013)
Crockett et al. (2014)
Jordan et al. (2015)
Chester & DeWall (2017)
Eadeh et al. (2017)
Osgood (2017)

These studies investigate the relation between punishment decisions and the punishers' mental and affective states but are agnostic about the feelings and beliefs of transgressors.

However, we argue that punishment is, to a large extent, driven by the punisher's consideration for the transgressor's affective and mental states.

THE THREE MOTIVES BEHIND PUNISHMENT



DISTRIBUTIVE JUSTICE

PREFERENCE OVER MATERIAL STATES:
THE DESIRE TO REDUCE THE
TRANSGRESSOR'S WELFARE



RETRIBUTIVE JUSTICE

PREFERENCE OVER AFFECTIVE STATES:
THE DESIRE TO MAKE THE
TRANSGRESSOR SUFFER



BELIEF-BASED MOTIVES

PREFERENCE OVER COGNITIVE STATES:
THE DESIRE TO MAKE THE
TRANSGRESSOR UNDERSTAND

THE CHALLENGE

Previous studies were unable to distinguish **belief-based motives** from **material** and **affective** considerations, since beliefs were perfectly correlated with the latter: the transgressors were aware whether, and why, they had been punished, and suffered accordingly.

E.g., in classic lab studies both parties know that there is a possibility of punishment or can infer from the final payments if they have been punished.



This project: **disentangle the three motives** by gradually introducing material, affective, and belief-based motives across different conditions.

EXPERIMENT

Prolific, $N = 1806$ (903 pairs)

Preregistered at [AsPredicted.org](https://aspredicted.org/#29436) (#29436)

Real-time online interaction between two participants (**SMARTRIQS**)

Real effort allocation (Stage 1) & real monetary consequences (Stage 2)

EXPERIMENT – STAGE I

Two participants (A & B) work on a boring “slider task”:



A & B have to complete 50 sliders combined for a **FIXED** compensation (\$1.50 each)

A chooses how to allocate the work:

- Option 1. A: 10 sliders B: 40 sliders
- Option 2. A: 5 sliders B: 45 sliders



B is informed about A’s decision,
but doesn’t see A’s choice set

EXPERIMENT – STAGE 2 (MAIN)

B is informed about A's (unfair) allocation of work

Both A & B are about to receive a *surprise* bonus of \$1

B makes a (private) decision before A is informed about the bonus:

- Do not reduce A's surprise bonus (“**no punishment**”)
- Reduce A's surprise bonus by \$0.50 (“**moderate punishment**”)
- Reduce A's surprise bonus by \$0.90 (“**severe punishment**”)

CONDITIONS

Four conditions (between subjects):

“IGNORANCE”

“SUFFERING”

“JUSTICE”

“REVENGE”

In ALL conditions, if B chooses the **no punishment** or **severe punishment**, A simply receives \$0.10 or \$1, without receiving any additional information about why they received this bonus, or even knowing what the maximum bonus was:

Do not reduce my partner's bonus.

Your partner will not receive any message.
(They will simply receive \$1.00.)



no punishment

Reduce my partner's bonus by \$0.90.

Your partner will not receive any message.
(They will simply receive \$0.10. They will
NOT KNOW that their bonus has been
reduced.)



severe punishment

CONDITIONS

However, if the participant chooses the **moderate punishment**, A also receives a message, which differs across conditions:

IGNORANCE

Reduce my partner's bonus by \$0.50.

Your partner will not receive any message.
(They will simply receive \$0.50. They will **NOT KNOW** that their bonus has been reduced.)



SUFFERING

Reduce my partner's bonus by \$0.50.

Your partner will receive the following message (They **WILL KNOW** that their bonus has been reduced, but they will **NOT KNOW** why their bonus has been reduced):

Your bonus has been reduced by 0.50.



JUSTICE

Reduce my partner's bonus by \$0.50.

Your partner will receive the following message (They **WILL KNOW** why their bonus has been reduced, but they will **NOT KNOW** who has reduced their bonus):

Your bonus has been reduced by 0.50, because you were unfair to your partner in the previous task.



REVENGE

Reduce my partner's bonus by \$0.50.

Your partner will receive the following message (They **WILL KNOW** why their bonus has been reduced, and they **WILL KNOW** who has reduced their bonus):

Your bonus has been reduced by 0.50. Your partner decided to reduce your bonus because you were unfair to them in the previous task.



HYPOTHESES



Just desserts hypothesis: no difference across conditions.



Comparative suffering hypothesis: more people choose the moderate punishment in the SUFFERING, JUSTICE, and REVENGE conditions than in the IGNORANCE condition, but there is no difference between the former three.



Understanding hypothesis #1: more people choose the moderate punishment in the JUSTICE and REVENGE conditions than in the IGNORANCE and SUFFERING conditions.



Understanding hypothesis #2: more people choose the moderate punishment in the REVENGE condition than in the JUSTICE condition.

RATIONALE FOR INCLUDING TWO FIXED OPTIONS (NO & SEVERE PUNISHMENT)

Investigate whether:

1. Participants punish MORE LIKELY when they can send a message (extensive margin)

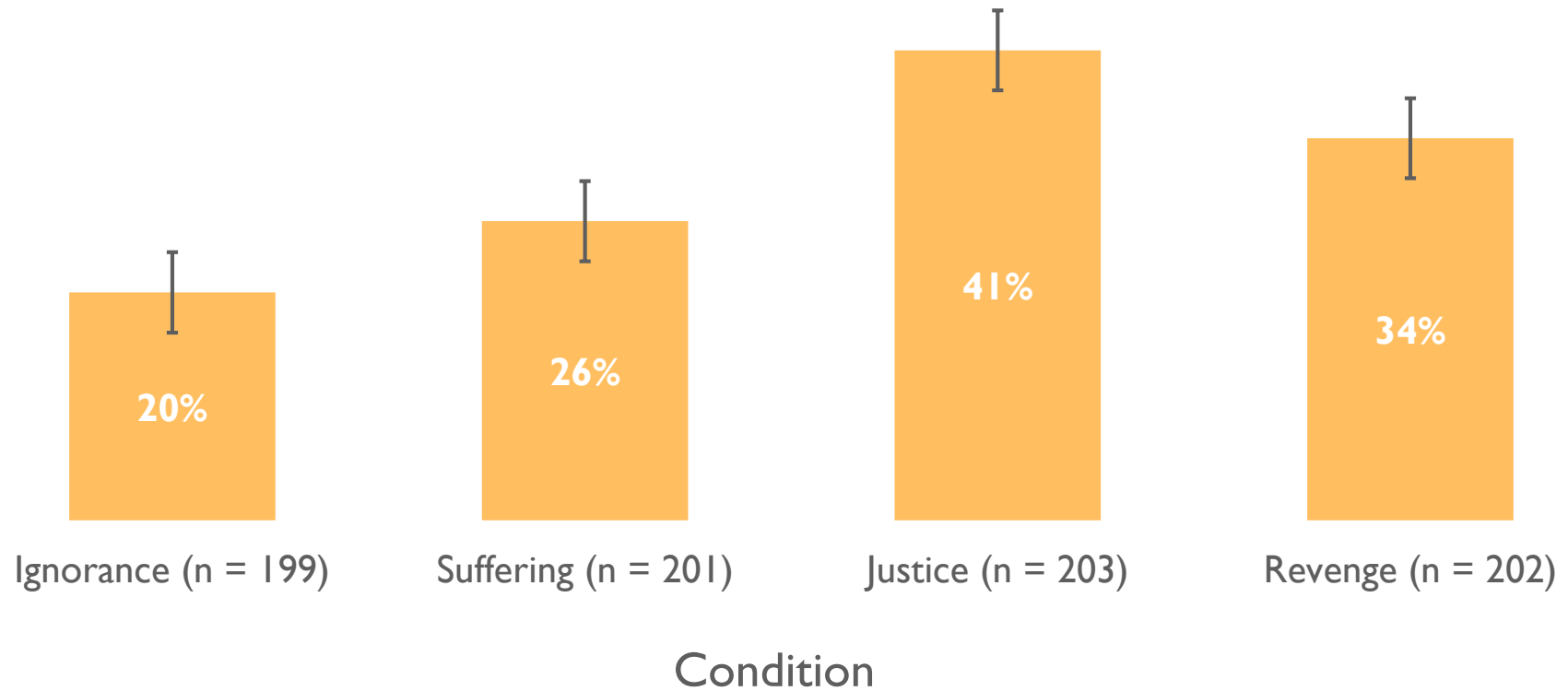
Pattern: % **no punishment** decreases and % **moderate punishment** increases

2. Participants punish LESS SEVERELY when they can send a message (tradeoff between distributive justice and other motives)

Pattern: % **severe punishment** decreases and % **moderate punishment** increases

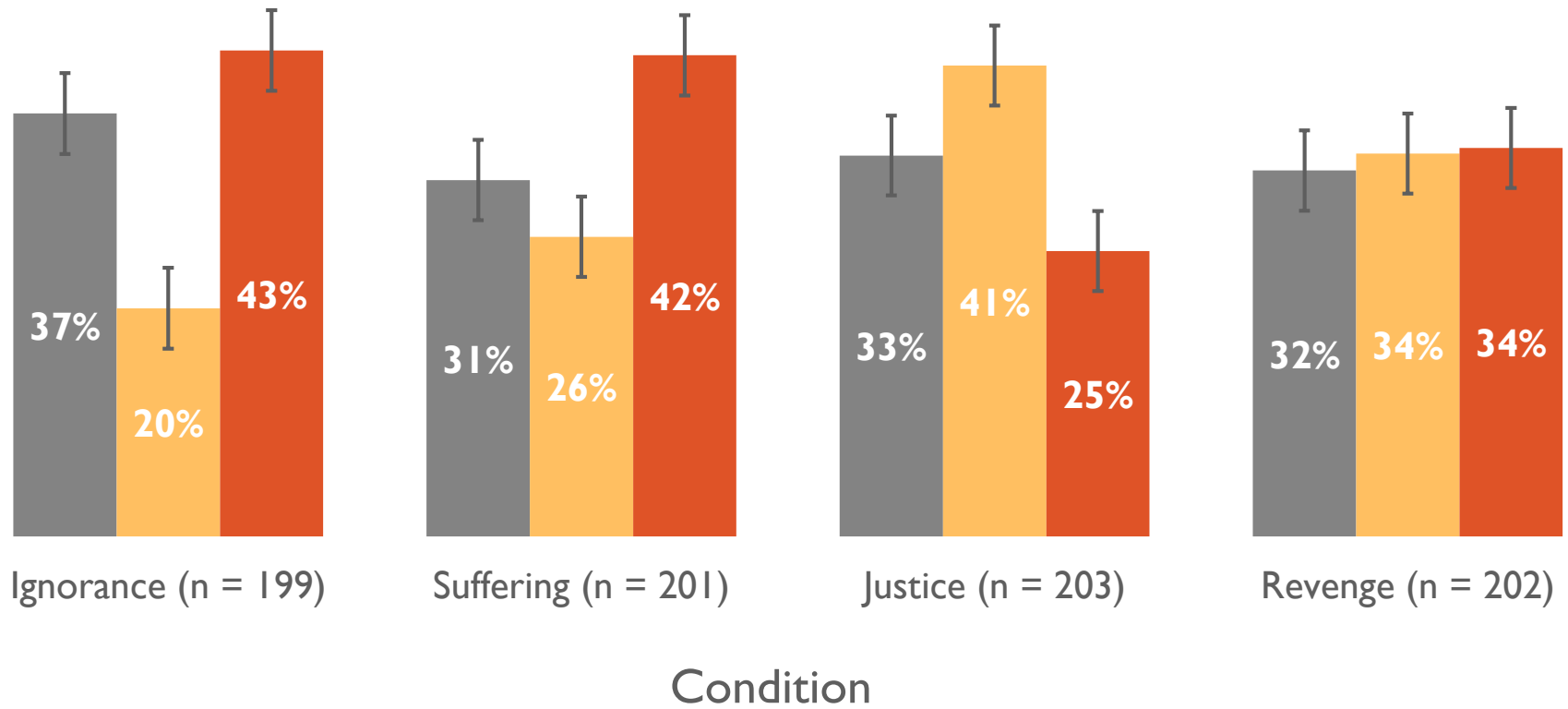
PROPORTION OF PARTICIPANTS CHOOSING MODERATE P. (± 1 SE)

■ MODERATE (-\$0.50) + message



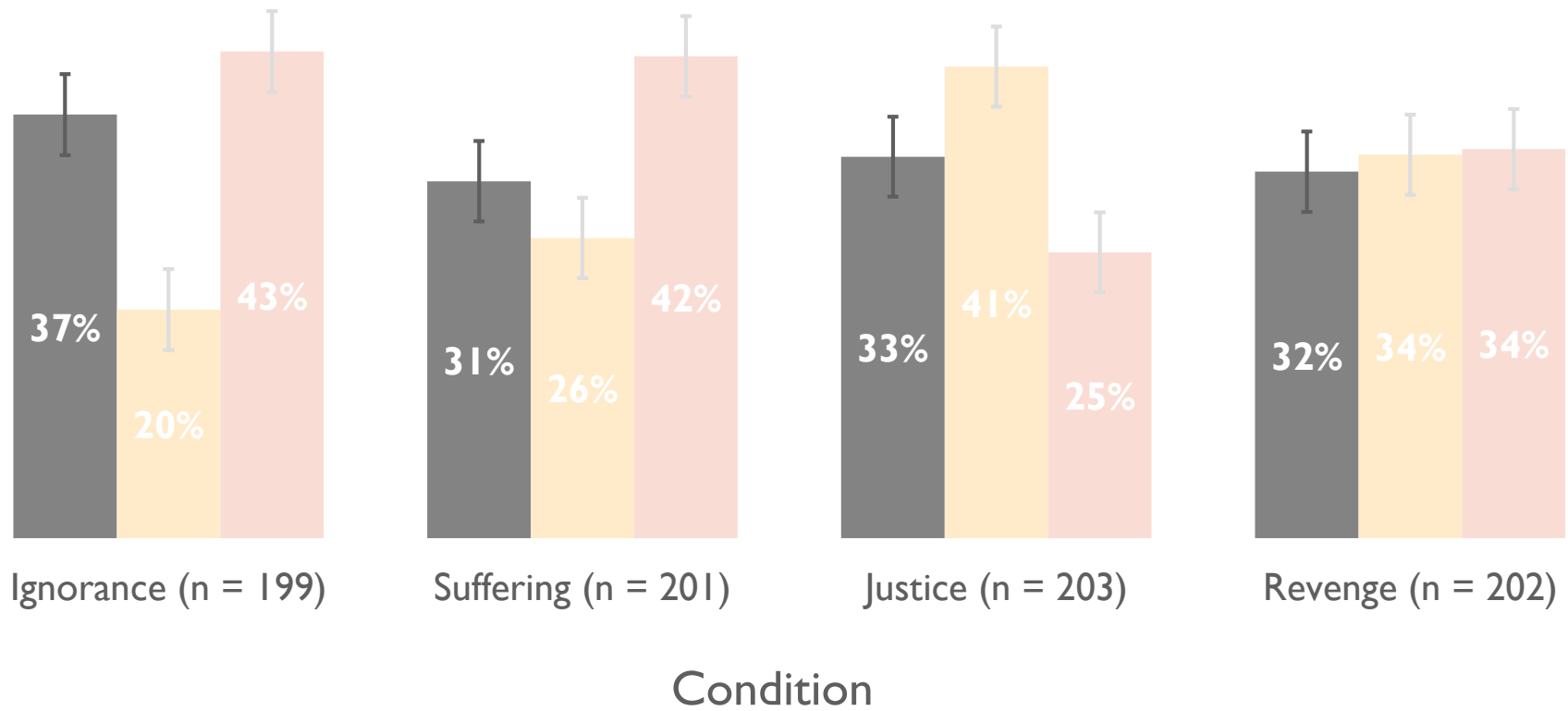
PROPORTION OF PARTICIPANTS CHOOSING EACH ACTION (± 1 SE)

■ NO (-\$0) ■ MODERATE (-\$0.50) + message ■ SEVERE (-\$0.90)



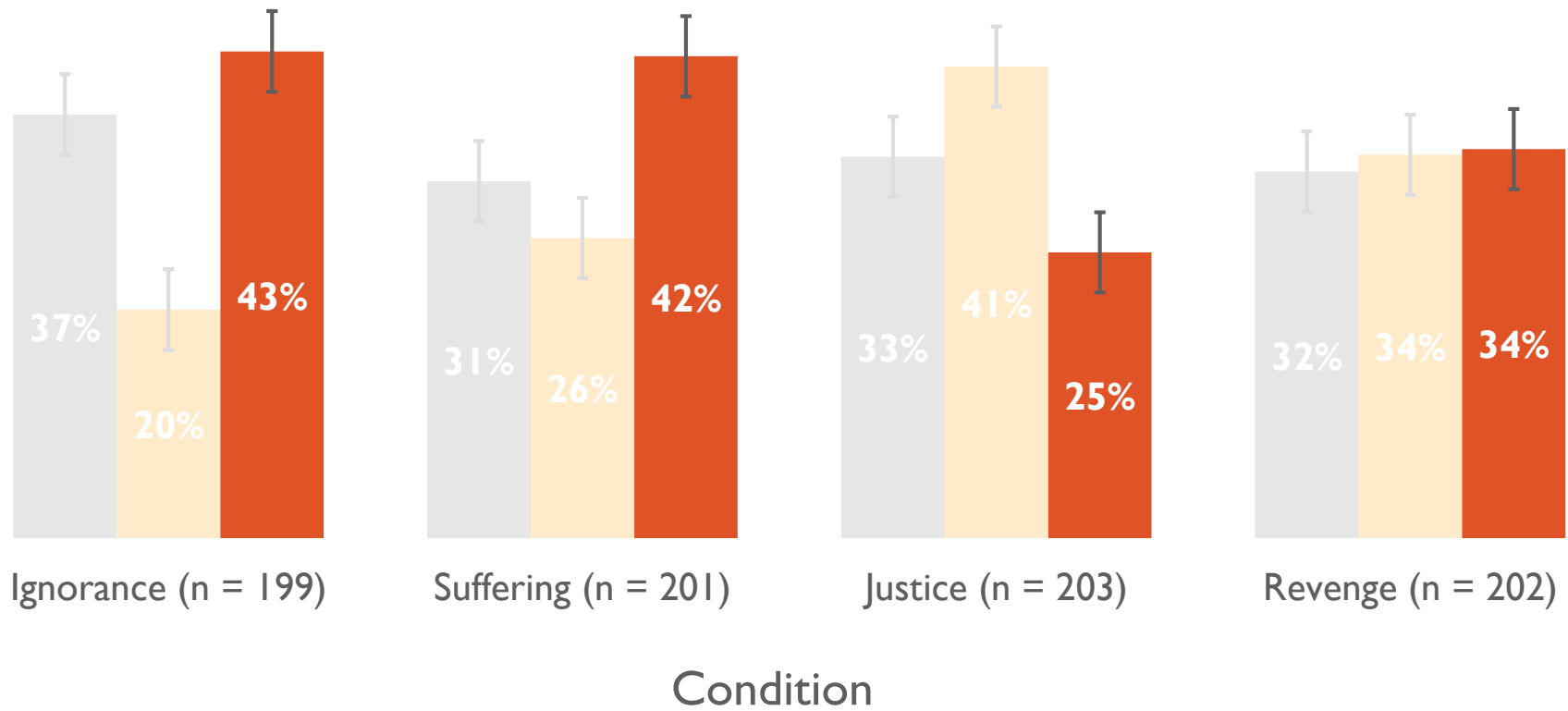
PROPORTION OF PARTICIPANTS CHOOSING EACH ACTION (± 1 SE)

■ NO (-\$0) ■ MODERATE (-\$0.50) + message ■ SEVERE (-\$0.90)



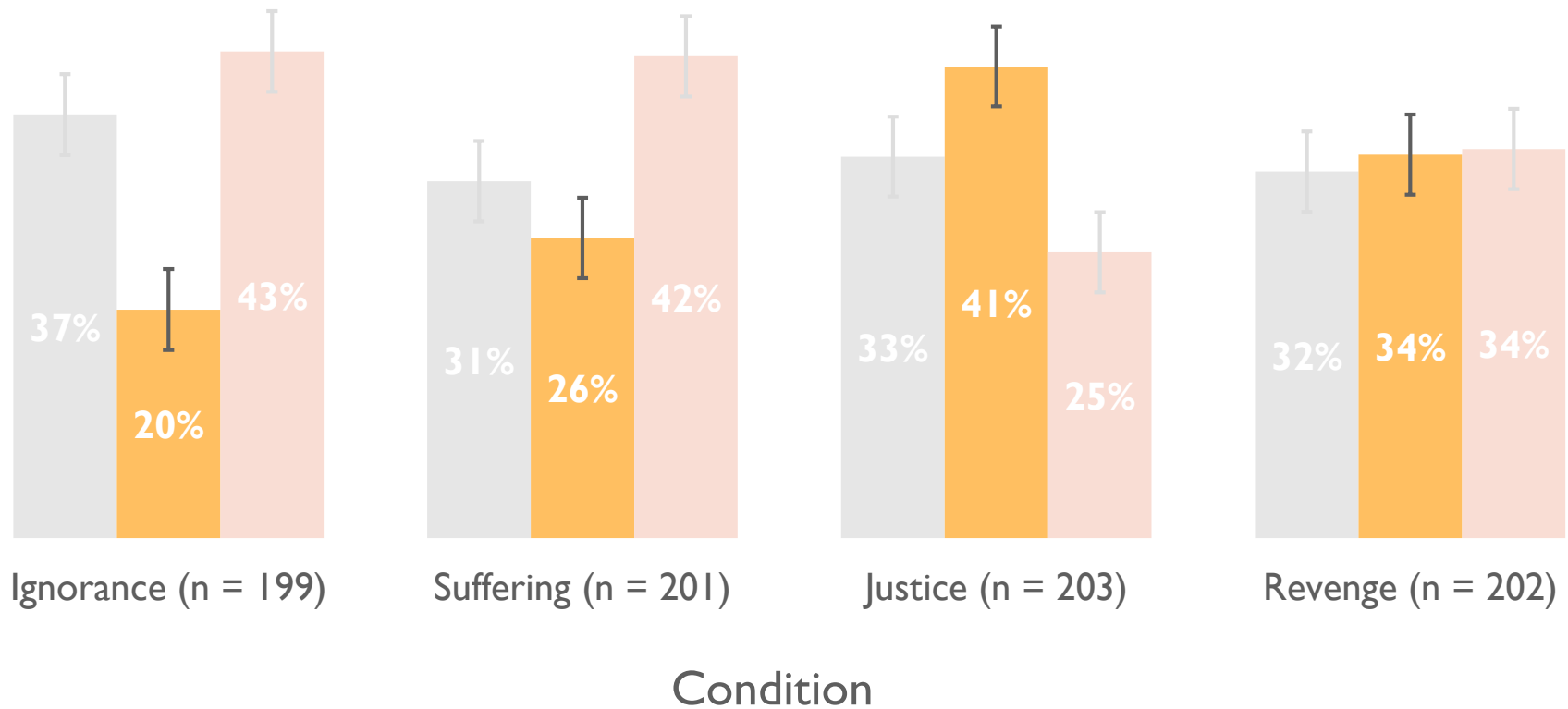
PROPORTION OF PARTICIPANTS CHOOSING EACH ACTION (± 1 SE)

■ NO (-\$0) ■ MODERATE (-\$0.50) + message ■ SEVERE (-\$0.90)



PROPORTION OF PARTICIPANTS CHOOSING EACH ACTION (± 1 SE)

■ NO (-\$0) ■ MODERATE (-\$0.50) + message ■ SEVERE (-\$0.90)





**PERSONAL
REVENGE**

OR



**IMPERSONAL
JUSTICE**

By making the punishment personal (justice condition → revenge condition), we introduced two potentially conflicting set of motives:

(1) Some participants want to let the transgressor know who punished them

- make sure that the transgressor fully understands
- signal agency and credibility
- standing up for themselves (honor culture)

(2) Other participants prefer to remain anonymous and seem impartial

- do not want to be perceived as petty or hot-headed
- do not want to look like a vigilante / violating social norms
- fear of retaliation

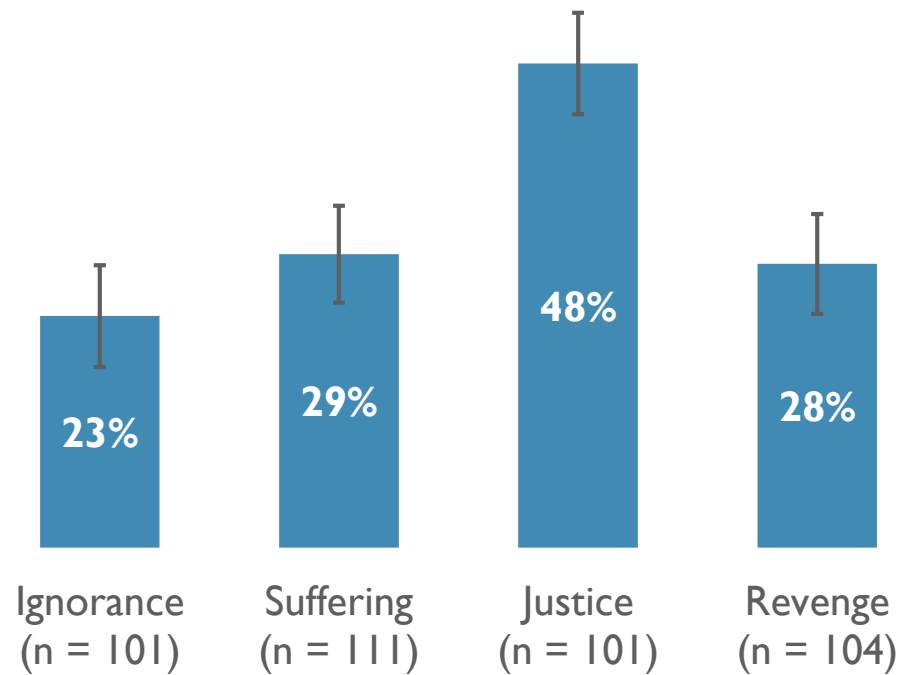
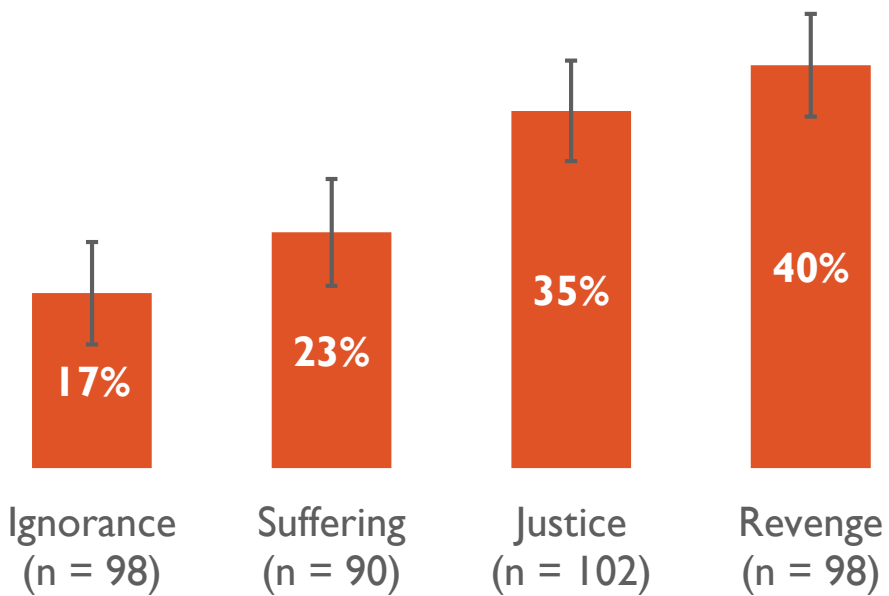


PERSONAL REVENGE

OR

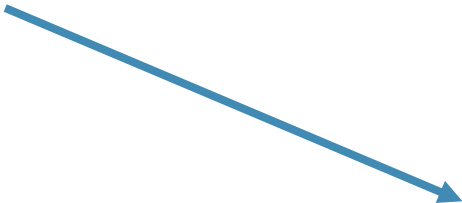


IMPERSONAL JUSTICE



IS THIS SIMPLY DETERRENCE / TEACHING A LESSON?

Would your partner treat others
WORSE or would your partner treat
others BETTER in the future?
[if you chose the moderate punishment]
(-100 ... +100)



4: Regression results: Likelihood of choosing MODERATE punishment

	<i>Dependent variable:</i>		
	moderate_pun_dummy		
	(1)	(2)	(3)
Suffer? (1 = y)	0.063 (0.045)	0.024 (0.045)	0.020 (0.045)
Explain? (1 = y)	0.150*** (0.045)	0.124*** (0.045)	0.125*** (0.045)
Identity? (1 = y)	-0.077* (0.045)	-0.075* (0.045)	-0.075* (0.044)
Behave better in future		0.002*** (0.0004)	0.002*** (0.0004)
Sex (1 = f)			0.032 (0.032)
Age (years)			-0.002* (0.001)
Constant	0.201*** (0.032)	0.210*** (0.032)	0.271*** (0.056)
Observations	805	805	805
R ²	0.030	0.058	0.062
Adjusted R ²	0.026	0.053	0.055

Dummy coding:

Suffer = 1:
suffering, justice, revenge

Explain = 1:
justice, revenge

Identity = 1:
revenge

Note:

*p<0.1; **p<0.05; ***p<0.01

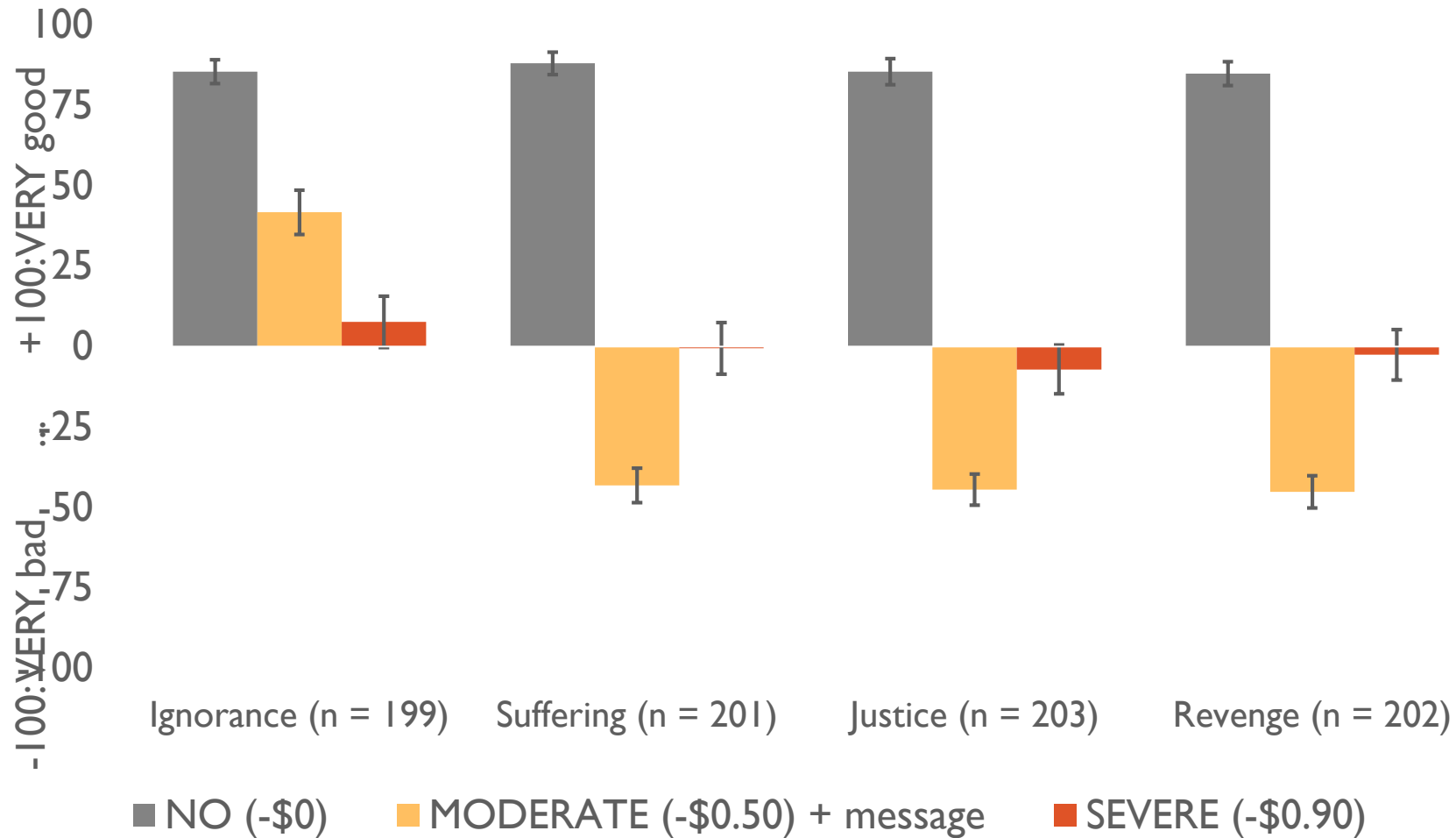
SUMMARY

We argue that **belief-based preferences play a crucial role in punishment decisions**, independently of distributive, retributive, and deterrent motives.

In a novel experiment we demonstrate that this **desire to affect beliefs is often prioritized over distributive and retributive preferences**: people who would otherwise enact harsh punishments, are willing to punish less severely, if by doing so they can tell the transgressor why they are punishing them.

MANIPULATION CHECK: SUFFERING

"Would YOUR PARTNER feel bad (experience suffering) or feel good (experience joy)?" (error bars: 95% CIs)



THE MORAL CHOICE

"How MORAL or IMMORAL would it be to choose the following options?" (error bars: 95% CIs)

