The Dice are Cast: The Role of Intended versus Actual Contributions in Responsibility Attribution

Tobias Gerstenberg, David A. Lagnado, Yaakov Kareev

contact: t.gerstenberg@ucl.ac.uk

UCL

"Even a dog can tell the difference between being stumbled over and kicked at."   - Justice Holmes

Can you?

Abstract

- How much are people's responsibility attributions affected by intentions versus outcomes in group contexts?
- Experimental game with noisy relationship between intentions and outcomes: good intentions can lead to bad outcomes
- Group's outcome depends on each individual's contribution → social dilemma
- Participants = independent judges, attributing blame/credit to players in a group

Research Question

$r \neq 1$

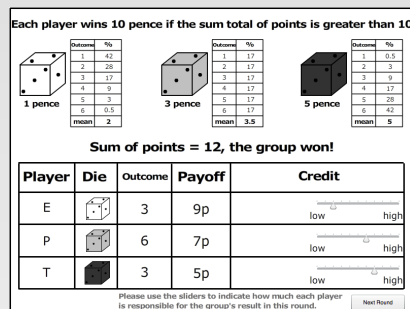**Intentions vs. Outcomes**

Responsibility

Hypotheses
1. Main effect of intention
2. Main effect of outcome
3. Outcomes weigh heavier than intentions[1]

Experiment

Rationale: Create noisy environment in which valence of intentions and outcomes can dissociate
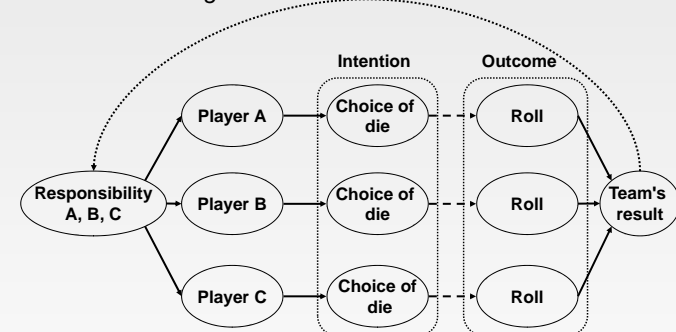
Task: Attribute credit/blame to players

Screenshot of the game

Details:
- team wins if sum > 10
- probability of winning:
  black die (68%) > grey die (50%) > white die (33%)
- expected payoff:
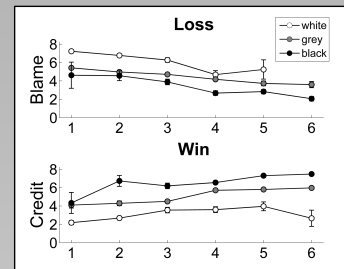 black die (1.8p) < grey die (2p) < white die (2.3p)

Underlying structure of of the game

1st part: 20 rounds, random choice of die, probability distribution of each die determines outcome

2nd part: 14 rounds with test cases (intentions and outcomes uncorrelated)

Results

1) Mean Responsibility Ratings

Aggregate level:
- main effect of choice of die (= intention)
- main effect of roll (= outcome)

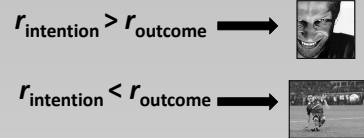**Are participants weighting both intentions and outcomes?**

2) Regression Analyses

intention-based model: $\beta_0 + \beta_{die}$
outcome-based model: $\beta_0 + \beta_{roll}$
mixture model: $\beta_0 + \beta_{die} + \beta_{roll}$

Model	R^2	F	β	t
intention	.268	1757*	.518[a]	41.93*
outcome	.219	1346*	.468[b]	36.69*
mixed	.303	1042*	.370[a]	24.02*
			.238[b]	15.48*

[a] = die, [b] = roll, * = p < .01

Classification

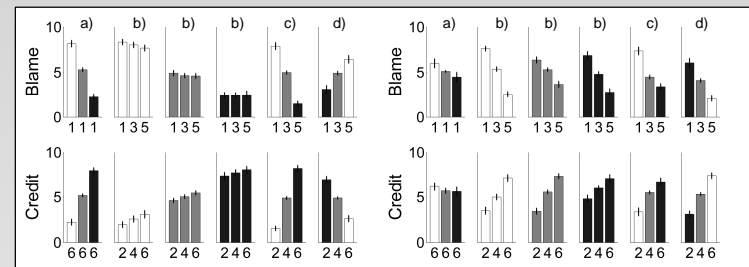$r_{intention} > r_{outcome}$ ⟶

$r_{intention} < r_{outcome}$ ⟶

How do intention-based vs. outcome based participants attribute responsibility on the test cases?

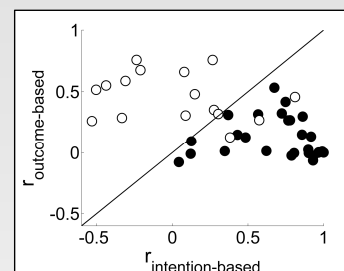### 3) Test Cases

| intention group (*N* = 29) | outcome group (*N* = 16) |

a) same outcome, different dice   c) congruent
b) same die, different outcome   d) incongruent

- most participants' responsibility attributions correlate higher with intention vs. outcome model
- only few participants seem to weigh both factors

● intention-based
○ outcome-based

Discussion

- majority of participants focused on intentions
- previously reported outcome bias weaker when participants acted as independent judges[1]
- possible explanations for individual differences:
  1) causal vs. moral interpretation of responsibility
  2) mentalizing
  3) illusion of control

References

Gerstenberg, T., Lagnado, D. A. & Kareev, Y. (2010). The dice are cast: The role of intended versus actual contributions in responsibility attribution. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

[1]Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a trembling hand game. *Plos One, 5, 1–7.*