The Influence of Outcome Severity on Ascriptions of Intention & Punishment

Aryn Pyke (apyke@connect.carleton.ca)
Deepthi Kamawar (dkamawar@ccs.carleton.ca)
Diana Ridgeway (dianar@davidridgeway.com)
Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive
Ottawa, ON, K1S 5B6 Canada

Keywords: intention; TOM; folk psychology; moral judgment

Introduction

Are adults' and childrens' ascriptions of intention and allocations of punishment affected not only by the valence of a protagonist's intention (positive/negative), but also by the undesirability of (potentially unintended) outcome(s) of the protagonist's action (e.g., something gets broken)? Even young children are sensitive to whether an outcome was intended when allocating blame/punishment (Nunez & Harris, 1998). However, the moral acceptability of the (possibly unintended) outcome of an action may also 'retroactively' influence the degree to which the outcome is judged intentional and thus blame/praiseworthy. In particular, negative unintended outcomes are more likely to be classified as intentional than positive unintended outcomes by children and adults (Knobe, 2003; Leslie, Knobe, & Cohen, 2006). In the present study, we focus on the negative side of the outcome spectrum (neutral, mildly negative, moderately negative), and we varied whether or not the protagonist's motive had been to achieve a negative or positive outcome. We discuss how intention valence and outcome severity influenced ascriptions of intention and punishment, and moral ratings of the protagonist.

Experiment

Method. Forty undergraduates were divided into 2 groups (N_A=18, N_B=22). Each read 6 stories which featured either a well-intentioned protagonist (group A) or a negatively-intentioned protagonist (group B). For each story there were 3 possible outcomes. For example, in the positive-intention version of one story, Sally wants to share her cookies but accidentally drops them, resulting in: none breaking (neutral outcome); 1 breaking (mildly negative outcome); or 8 breaking (moderately negative outcome). In the negative-intentioned version of the story, Sally deliberately throws the cookies on the floor to break them and avoid sharing -- again the same three possible outcomes apply (neutral, mildly negative, and moderately negative). The outcomes varied across stories with each participant receiving two per type. Participants then ascribed intention ('Did she mean to drop the cookies?'), gave a moral rating for the protagonist (5-point scale), and assigned punishment (0=no punishment, 1=a little trouble, 2=a lot of trouble).

Results. 2 (intention: positive, negative) x 3 (outcome: neutral, mildly negative, moderately negative) ANOVAs were conducted for each dependent variable: ascription of intention, moral rating, and assigned punishment.

Ascription of Intention. Participants recognized that positively-intentioned protagonists didn't "mean to" cause the (negative) outcome, whereas negative-intentioned protagonists did, $F(1,38)=342.4$, p=.000. The severity of the outcome did not influence whether the outcome was classified as intentional, $F(1.65,62.52)=1.9$, p=.161.

Moral Rating. The protagonist's intention (pos, neg) influenced participants' moral ratings for the protagonist, $F(1,38)=126.6$, p=.000. For each outcome, moral ratings were higher when the protagonist's intention was positive than negative. However, negativity of the outcome also influenced moral ratings for the protagonist, $F(1.54,76)=14.6$, p=.000. Even for positive-intentioned protagonists, neutral outcomes prompted higher moral ratings than mild (p=.052) and moderate (p=.000) outcomes, and mild outcomes prompted higher ratings than moderate outcomes (p=.022). There was no interaction between intention and outcome severity, $F(2,76) = 1.9$, p = .153

Punishment. Degree of punishment was influenced both by intention valence and outcome severity, $F(2,76)=15.8$, p=.000, however there was no interaction. For both positive- and negative-intentioned protagonists, less punishment was assigned when the outcome was neutral versus mildly or moderately negative (ps<=.001), but there was no difference in assigned punishment between the latter two (p =.561).

Discussion. Participants correctly discriminated whether the protagonists brought about the outcome by accident or on purpose, but moral ratings for the protagonist were nonetheless coloured by the severity of outcome. Further, a negative intention was sufficient to warrant some assignment of punishment, even if the action did not succeed in producing a negative outcome (no breakage). A positive intention was not sufficient to avoid allocation of punishment. More punishment was allocated when the outcome involved some destruction of property than when it did not, however degree of punishment did not vary further according to whether the degree of destruction was mild or moderate (whether some vs. all of the cookies broke).

References

Knobe, J. (2003). Intentional action in folk psychology. *Philosophical Psychology, 16,* 309-324.

Leslie, A., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect. *Psychological Science, 17,* 421-427.

Nunez, M. & Harris, P. L. (1998). Psychological and deontic concepts. *Mind & Language, 13,* 153-170.