Neuropsychologia 46 (2008) 2949–2957

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

The influence of prior record on moral judgment☆

Dorit Kliemann[a,b,*], Liane Young[b], Jonathan Scholz[b], Rebecca Saxe[b]

[a] Department of Neuropsychology and Behavioral Neurobiology, Center for Cognitive Sciences, Hochschulring 18, Cognium Building, D-28359 Bremen, Germany
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 30 January 2008
Received in revised form 29 April 2008
Accepted 5 June 2008
Available online 20 June 2008

Keywords:
Theory of Mind
Moral reasoning
Temporo-parietal junction
Economic game
fMRI

ABSTRACT

Repeat offenders are commonly given more severe sentences than first-time offenders for the same violations. Though this practice makes intuitive sense, the theory behind escalating penalties is disputed in both legal and economic theories. Here we investigate folk intuitions concerning the moral and intentional status of actions performed by people with positive versus negative prior records. We hypothesized that prior record would modulate both moral judgment and mental state reasoning. Subjects first engaged in an economic game with fair (positive prior record) and unfair (negative prior record) competitors and then read descriptions of their competitors' actions that resulted in either positive or negative outcomes. The descriptions left the competitors' mental states unstated. We found that subjects judged actions producing negative outcomes as more "intentional" and more "blameworthy" when performed by unfair competitors. Although explicit mental state evaluation was not required, moral judgments in this case were accompanied by increased activation in brain regions associated with mental state reasoning, including predominantly the right temporo-parietal junction (RTPJ). The magnitude of RTPJ activation was correlated with individual subjects' behavioural responses to unfair play in the game. These results thus provide insight for both legal theory and moral psychology.

1. Introduction

Repeat offenders commonly receive more severe sentences than first-time offenders for the same violations. This principle of escalating penalties with offense history is widespread in both criminal and civil law, in many countries and over many centuries (Durham, 1987). Moreover, the practice fits with common sense: intuitively, it seems "right" that persistent offenders receive more severe punishments. Nevertheless, both justice and economic models of the law advocate against escalating penalties. According to the justice model, punishment is justified only if the amount of punishment is proportional to the harm caused by the violation. Escalating penalties violate this rule, punishing repeat offenses disproportionately (Ashworth, 2005; Durham, 1987). According to the economic model, an optimal punishment regime is one in which the expected punishment for a violation equals the social cost of the violation. Expected punishment is a function of both the penalty once caught, and the probability of being caught. Since repeat offenders are more likely to be caught than first-time offenders, their expected punishment escalates even if the amount of the penalty does not (Dana, 2001; Emons, 2007).

In spite of these considerations, legal practice in the US over the past 30 years has tended towards increasing, rather than decreasing, reliance on prior record during sentencing, as in the "Three-Strikes" policy in California (Austin, Clark, Hardyman, & Henry, 2000). Many efforts have been made to account for this phenomenon (Ashworth, 2005; Dana, 2001). One theory, for example, treats escalating penalties as deterrence or preventative incapacitation: if the offender is incarcerated, he or she will be less able to commit another offense (Ashworth, 2005).

An alternative is that escalating penalties express society's moral condemnation of persistent wrongful action (Dana, 2001; Sunstein, 2005), regardless of utilitarian calculations. The current study investigates this alternative: do laypersons indeed judge first-time offenders as less blameworthy, and repeat offenders as more blameworthy, for the same harm caused? How are the effects of prior record related to other aspects of folk morality, such as attribution of intent to moral agents (Cushman, personal communication; Pizarro, Laney, Morris, & Loftus, 2006; Woolfolk, Doris, & Darley, 2006)? Specifically, does negative prior record lead subjects to attribute more intentionality to agents for causing negative outcomes; if so, is this effect a cause or a consequence of a change in moral judgment, i.e. increase in blame.

Consider the following example scenario: *Ashley works at the computer help desk and often friends bring their computers. Once,*

☆ The study was carried out at the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

* Corresponding author at: Max-Planck-Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Tel.: +49 30 824 06 473.
E-mail address: kliemann@mpib-berlin.mpg.de (D. Kliemann).

0028-3932/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2008.06.010

Table 1
Story Task vignettes

Negative outcome story	Target sentence	Positive outcome story	Target sentence
Jessica once went on a camping trip with her ex-boyfriend. On the second day, there was a thunderstorm, and a big branch fell onto the tent, hitting him on the ankle. She wrapped an ace bandage tightly around the swelling ankle, which made the swelling get worse, and the pain even more intense	Jessica made her ex-boyfriend's swelling ankle worse when she wrapped it	Jessica once went on a camping trip with her ex-boyfriend. On the second day, there was a thunderstorm, and a big branch fell onto the tent, hitting him on the ankle. She wrapped the swelling ankle in a sheet that was soaked from the cold rainwater. The cold water numbed the pain and helped him recover	Jessica wrapped her ex-boyfriend's swelling ankle and it helped him to recover
Chris found someone else's clothes lying wet in the washing machine in the basement of his building. He put all of the clothes into the dryer and turned it on the regular cycle, shrinking his neighbor's new sweater four sizes	Chris shrank his neighbor's new sweater	Chris was doing his laundry very late at night, in the basement of his building. Mixed in with his own dry clothes were someone else's clothes. He kept folding until all the clothes were done: his own, and the stranger's	Chris folded some of the stranger's dry clothes

In each scenario, one of the ten competitors from the Game performed an action that either lead to a positive or a negative outcome. Scenarios did not explicitly state the agent's intentions or the action's moral status. Corresponding to each story, a target sentence was presented to ask for subject's rating of the intentional (Behavioural Experiment) or moral status (fMRI Experiment) of the action.

her ex-boyfriend brought his computer, which had crashed. Ashley restarted the computer, the hard-drive was re-formatted and all of Chris' files were lost. For actions resulting in a negative outcome (e.g., lost files on the computer) caused by an agent with a negative prior record (e.g., a negative prior personal experience with Ashley), we hypothesize that participants judge the agent as (1) more blameworthy and (2) having acted more intentionally, compared to agents with no prior record. If so, we further ask whether the increase in blame precedes or follows the increased attribution of intentionality. The current study investigated these questions, using behavioural and neuroimaging (functional magnetic resonance imaging, fMRI) methods.

Subjects read a series of short vignettes about an agent's action and the subsequent positive or negative outcome (Story Task). The stories left the mental states (e.g., thoughts, desires, intentions) of the agents unstated (for an example of the vignettes see Table 1), making both the moral and intentional status of the actions ambiguous. Subjects then judged the intentional status of the actions (Behavioural Experiment), or the moral status of the actions (fMRI Experiment).

To manipulate the perceived "prior record" of the agents in the stories, the subjects were exposed to a (purportedly) real social interaction (the Game) with the same agents prior to participating in the Story Task. The social interaction took place in the context of an economic game; fairness and trustworthiness are emotionally salient and morally valenced features of social behaviour that can be manipulated realistically in the lab (Berg, Dickhaut, & McCabe, 1995; Haselhuhn & Mellers, 2005; Koenigs & Tranel, 2007; Rabin, 1993; Singer, Kiebel et al., 2004; Singer et al., 2006). Subjects played against 10 competitors; half of the competitors played fairly (positive prior record), and the others played unfairly (negative prior record). We then assessed the influence of prior record on subjects' subsequent judgments about the competitors' actions in the Story Task.

In particular, we investigated the patterns, and neural correlates, of folk intuitions about the intentional status of repeat versus first-time offenses. To this end, subjects in the fMRI Experiment also performed a second task while in the scanner, designed to identify brain regions previously implicated in mental state reasoning or Theory of Mind (Baron-Cohen, Leslie, & Frith, 1985; Flavell, 1999; Leslie, Friedman, & German, 2004; Premack & Woodruff, 1978; Saxe, 2006; Saxe, Carey, & Kanwisher, 2004). Previous research on the neural basis of Theory of Mind has identified a consistent group of brain regions recruited when participants reason about another agent's beliefs, desires, and/or intentions: the temporo-parietal junction (bilaterally) (RTPJ, LTPJ), the precuneus (PC) and the medial prefrontal cortex (MPFC) (Fletcher et al., 1995; Gallagher et al., 2000; Ruby & Decety, 2003; Saxe & Kanwisher, 2003; Vogeley et al., 2001). Of these regions, the RTPJ appears to be the most selective for belief attribution (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Powell, 2006; Saxe & Wexler, 2005). We therefore hypothesized that the response profile in these brain regions, especially, the RTPJ, would provide evidence concerning the influence of prior record on mental state reasoning during moral judgment.

2. Methods

Subjects (fMRI Experiment: nine male, seventeen female, aged 19–33 years; Behavioural Experiment: three male, four female, aged 18–48 years) were naïve to experimental hypotheses, right-handed and recruited by email at the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT). All subjects had normal or corrected-to-normal vision, were native English speakers, participated for payment and gave written informed consent in accordance with the requirements of MIT's Committee on the Use of Humans as Experimental Subjects. Each subject participated two sessions: the Game, and the Story Task.

2.1. Game

Subjects were told that they were recruited in groups of 10–12 people, who would not meet face to face, but would play the Game against each other over a computer network. When they arrived, subjects were met by an experimenter and taken to a hallway containing 12 experimental rooms, each labeled with the experiment name, and a subject number. Subjects were taken into one room containing a single computer and a sheet of paper, their photograph was taken, and they were given written instructions for the Game (described below). Subjects were informed that in the second experimental session, members of the group would read stories about one another. They were asked to provide two or three short stories that would be rewritten by the experimenters and later presented to the players. They were provided hints of possible story types (e.g., something nice you did for a stranger, something that turned out worse than expected).

After a few minutes, the experimenter returned, collected the stories, and showed the subject a page containing photographs of the 10 "other players". Subjects were asked to mark on the page whether they knew any of the people in the photographs. The photographs were taken from the FRI CVL database of face images (http://www.lrv.fri.uni-lj.si/facedb.html, Solina, Peer, Batagelj, Juvan, & Kovac, 2003) and showed six male and four female white college-aged faces. All subjects marked that they did not know any of the players. Then subjects were instructed to wait for a cue, on the screen, that everyone else was ready, and that the Game was about to begin. The experimenter then left the room, and within 2 min triggered the Game remotely.

Subjects then played 100 trials of repeated sequential economic investment game. For each trial, an Investor and a Trustee were chosen: the subject was ran-