SOCIAL NEUROSCIENCE, 2017
VOL. 12, NO. 3, 308–329
http://dx.doi.org/10.1080/17470919.2016.1175380

Routledge
Taylor & Francis Group

Elevated moral condemnation of third-party violations in multiple sclerosis patients

Indrajeet Patil[a], Liane Young[b], Vladimiro Sinay[c] and Ezequiel Gleichgerrcht[d]

[a]Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy; [b]Psychology Department, Boston College, Chestnut Hill, MA, USA; [c]Department of Neurology, Institute of Cognitive Neurology, Buenos Aires, Argentina; [d]Department of Neurology, Medical University of South Carolina, Charleston, SC, USA

ABSTRACT
Recent research has demonstrated impairments in social cognition associated with multiple sclerosis (MS). The present work asks whether these impairments are associated with atypical moral judgment. Specifically, we assessed whether MS patients are able to integrate information about intentions and outcomes for moral judgment (i.e., appropriateness and punishment judgments) in the case of third-party acts. We found a complex pattern of moral judgments in MS patients: although their moral judgments were comparable to controls' for specific types of acts (e.g., accidental or intentional harms), they nevertheless judged behaviors to be less appropriate and endorsed more severe punishment across the board, and they were also more likely to report that others' responses would be congruent with theirs. Further analyses suggested that elevated levels of externally oriented cognition in MS (due to co-occurring alexithymia) explain these effects. Additionally, we found that the distinction between appropriateness and punishment judgments, whereby harmful outcomes influence punishment judgments to a greater extent than appropriateness judgments, was preserved in MS despite the observed disruptions in the affective and motivational components of empathy. The current results inform the two-process model for intent-based moral judgments as well as possible strategies for improving the quality of life in MS patients.

ARTICLE HISTORY
Received 15 September 2015
Revised 30 March 2016
Published online
21 April 2016

KEYWORDS
Morality; punishment; empathy; alexithymia; multiple sclerosis; moral luck

1. Introduction

Multiple sclerosis (MS) is an autoimmune disease involving demyelination of nerve cells as well as neural inflammation and degeneration, which leads to a wide range of physical, cognitive, and psychiatric signs and symptoms. In addition to the study of physical impairments and neuropsychological dysfunction (Feinstein, Magalhaes, Richard, Audet, & Moore, 2014; Pepping, Brunings, & Goldberg, 2013; Rocca et al., 2015), recent research has also begun to unravel problems associated with social cognition in MS.

1.1. *MS and deficits in social cognition*

Social cognition is an umbrella term for the set of processes necessary for effectively dealing with social situations, including perception and evaluation of socio-emotional stimuli as well as regulation and modulation of one's social behavior according to such evaluations (Adolphs, 2009).

Studies in MS have shown impairments in the theory of mind (ToM), the capacity to represent others' epistemic states (i.e., thoughts, beliefs, desires, etc.; Koster-Hale & Saxe, 2013). In particular, MS patients perform poorly in extracting information about intentionality, representing first- and second-order false beliefs, and detecting social faux pas in verbal/nonverbal ToM assessment tasks (Banati et al., 2010; Charvet, Cleary, Vazquez, Belman, & Krupp, 2014; Henry et al., 2011; Pöttgen, Dziobek, Reh, Heesen, & Gold, 2013; Roca et al., 2014). These deficits remain even after controlling for the co-occurring neuropsychological impairments, such as executive function deficits (Charvet et al., 2014; Pöttgen et al., 2013). Importantly, ToM deficits appear from the very early stage of the disease in young adults (Kraemer et al., 2013) and are found even in pediatric-onset MS patients (Charvet et al., 2014). These ToM deficits are also known to get worse with the disease duration/physical disability (Banati et al., 2010) and cognitive impairment (Ouellet et al., 2010) in MS.

CONTACT Ezequiel Gleichgerrcht gleichge@musc.edu Department of Neurology, Medical University of South Carolina, 96 Jonathan Lucas St. Suite 301, Charleston, SC 29425, USA
Supplemental data for this article can be accessed here.

MS patients have also been shown to perform atypically on empathy tasks that index the cognitive ability to form representations of feeling states (pain, emotions, etc.) in others and affective sharing of these states (de Vignemont & Singer, 2006). Self-report questionnaires reveal disruptions in general empathizing skills in MS (Kraemer et al., 2013), with a more specific reduction in empathic concern (EC) alongside elevated levels of personal distress (PD) (Gleichgerrcht, Tomashitis, & Sinay, 2015). MS patients also show impaired performance on tasks requiring them to infer emotional states from facial (Henry et al., 2009), bodily (Cecchetto et al., 2014), or prosodic (Beatty, Orbelo, Sorocco, & Ross, 2003) affective cues, as well as an abnormal pattern of neural processing of emotional faces (Jehna et al., 2011; Krause et al., 2009).

Additionally, MS patients have been shown to exhibit elevated levels of alexithymia, which is a dimensional personality construct defined by difficulties in analyzing, identifying, and communicating emotional states and externally focused cognitive style (Nemiah, Freyberger, & Sifneos, 1976). Whereas the preponderance rate of clinical alexithymia is about 10% in the healthy population (Bird & Cook, 2013), it is found to vary from 13.8 to as high as 50% in the MS population. Alexithymia also reliably predicts elevated levels of fatigue, depression, and anxiety in MS (Bodini et al., 2008; Chahraoui, Duchene, Rollot, Bonin, & Moreau, 2014; Gay, Vrignaud, Garitte, & Meunier, 2010; Gleichgerrcht et al., 2015).

1.2. Intent-based moral judgments and its neural basis

The two-process model of intent-based moral judgments (Cushman, 2008, 2015; Cushman, Sheketoff, Wharton, & Carey, 2013) posits two distinct processes that guide moral judgment: (*i*) a causal reasoning process, active in the presence of a harmful outcome (e.g., victim suffering), that involves an analysis of the agent's causal role in producing the outcome ("causal responsibility = bad"); and (*ii*) an intent-based reasoning process that involves an analysis of the agent's intent to bring about the outcome, leading to condemnation in the presence of a culpable mental state ("malicious belief/desire/intent = bad"). Although the outputs of these processes need not conflict (i.e., in the case of neutral acts or intentional harms), these two processes may yield different judgments (i.e., in case of accidental or attempted harms). The ultimate judgment in the case of conflicting outputs is the result of a competitive interaction between antecedent evaluations, and informed by the relative weight (which itself is determined, in part, by underlying personality traits; Prehn et al., 2008) assigned to the output of each process (Buckholtz et al., 2015; Young, Cushman, Hauser, & Saxe, 2007).

1.2.1. Mental-state reasoning process

Abundant evidence shows that both older children and healthy adults rely primarily on mental state information when evaluating third-party harms (Alter, Kernochan, & Darley, 2007; Baird & Astington, 2004; Barrett et al., 2016; Cushman, 2008; Gummerum & Chu, 2014). Behaviorally, individuals tend to forgive accidental harms based on innocent intentions, while they condemn attempted harms based on malicious intentions despite the absence of harmful outcomes (Cushman, 2008). At the neural level, the rTPJ has been shown to be a key region within the ToM network for mental state attribution during moral judgment (for a review, see Young & Tsoi, 2013). The rTPJ is recruited most robustly in response to attempted harms, when the perpetrator intends but fails to harm someone, and thus the condemnation relies heavily on intent information (Gan et al., 2015; Young et al., 2007; Young & Saxe, 2008); in fact, disrupting activity in the rTPJ leads to reduced condemnation of attempted harms (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010). Additionally, neurological patients (Baez, Couto, et al., 2014; Baez, Manes, et al., 2016; Baez, Kanske, et al., 2014; Baez, Morales, et al., 2016; Ciaramelli, Braghittoni, & di Pellegrino, 2012; Young et al., 2010) and sadistic individuals (Trémolière & Djeriouat, 2016) deliver more favorable assessments of attempted harms due to a reduced emotional response to harmful intent. Note that in case of condemning attempted harms, the causal reasoning process remains silent in the absence of a harmful outcome, and the intent-based process dominates moral judgment (Cushman, 2008; Young et al., 2007).

Compared to attempted harms, accidental harms pose a more difficult case. Forgiving accidental harms requires a robust representation of innocent intent that can counteract a prepotent tendency to condemn the actor based on the harmful outcome (cf. Moran et al., 2011). Accordingly, individuals with a higher overall response in the rTPJ (Young & Saxe, 2009) and greater differentiation of intentional versus accidental harms in the spatial pattern of activity in the rTPJ (Chakroff et al., 2016; Koster-Hale, Saxe, Dungan, & Young, 2013) tend to forgive accidental harms more. This can happen as early as 62 ms post-stimulus (Decety & Cacioppo, 2012) and involves the down-regulation of emotional arousal encoded in the amygdala in response to harm (Hesse et al., 2016; Ngo et al., 2015; Treadway et al., 2014; Yu, Li, & Zhou, 2015). What is more, stimulating this patch of cortex increases the role that belief information plays in moral judgments (Ye et al., 2015), as evidenced by more lenient judgments of accidental harms (Sellaro et al., 2015). Meanwhile, imposing a cognitive load on participants, and thereby taxing integration of mental state information, leads to the opposite pattern (Buon, Jacob, Loissel, & Dupoux, 2013). Finally, populations with impaired mental state inference, such as autism spectrum disorder, show abnormal patterns in the rTPJ (i.e., lack of differentiation between intentional and accidental harms in the spatial patterns of activity in the rTPJ; Chakroff et al., 2016; Koster-Hale et al., 2013) and deliver harsher moral judgments of accidental harms (Buon, Dupoux, et al., 2013; Moran et al., 2011; Salvano-Pardieu et al., 2015; but see Baez et al., 2012).

1.2.2. Causal reasoning process

Causal analyses of harmful events begin with the detection of a harmful outcome and subsequent search for a causally responsible agent (Sloman, Fernbach, & Ewing, 2009). The degree to which individuals pay attention to the causal role of actors who accidentally bring about negative outcomes is in turn determined by the extent to which they empathize with the victim. In other words, understanding and feeling victim distress can motivate individuals to condemn accidental harm-doers more based on causal involvement. Evidence on individual differences in moral judgment is broadly consistent with this idea. For example, individuals who score high on self-report measures of dispositional empathy are more inclined to condemn accidental harms (Trémolière & Djeriouat, 2016, Study 1). In addition, individuals with a certain genetic variation of the oxytocin receptor gene that predisposes them to being more empathic are more reluctant to exculpate accidental harm-doers (Walter et al., 2012). Subclinical (e.g., alexithymia; Patil & Silani, 2014b) and clinical (e.g., psychopathy; Young, Koenigs, Kruepke, & Newman, 2012) personalities characterized by reduced EC for others also exhibit an increased tendency to forgive accidents, arguably because they are less motivated to hold the agent causally responsible in the absence of strong empathic aversion. Finally, while evaluating harmful situations, participants in general spend more time looking at the victim than the perpetrator and exhibit increased activity in the empathy network (Decety, Michalska, & Kinzler, 2012).

1.2.3. Differential effect of moral luck on punishment and wrongness

Importantly, according to the two-process model, information about intent and cause are recruited to different degrees for different types of moral judgments. In particular, blame/punishment judgments rely to a greater degree on outcome information as compared to wrongness/appropriateness judgments, which rely more on intent information (Cushman, 2008, 2015; Cushman, Dreber, Wang, & Costa, 2009; Cushman et al., 2013). For instance, although the behaviors of two drivers who fall asleep at the wheel while driving under the influence are judged to be equally *wrong* or *inappropriate*, the driver who runs over and kills someone is *punished* more severely than the driver who runs into a tree. This asymmetric reliance on outcomes for punishment judgments (vis-à-vis appropriateness) has convincingly been argued to be an upshot of the ultimate evolutionary function of punishment (Martin & Cushman, 2016), i.e., to utilize the learning capacity of social partners to modify harmful behavior, including unintentionally harmful behavior. At the mechanistic level, however, this approach is implemented via inflexible moral outrage towards the harm-doer (Martin & Cushman, 2016), possibly stemming from empathy with the victim.

To summarize, judgments of third-party moral acts rely on ToM skills for generating robust representations of the agent's mental states, irrespective of whether a harmful outcome is present. Empathic evaluation of victim suffering (in the presence of a harmful outcome) additionally contributes to moral evaluation. Additionally, this empathic aversion exerts greater influence on punishment judgments versus wrongness judgments. Finally, these underlying psychological mechanisms are neurally implemented in interactions between limbic and frontotemporal structures.

1.3. Past work and current study

Given the aforementioned sociocognitive and socioaffective deficits in the MS population, particularly in ToM and empathy skills, we expect MS patients to exhibit irregular patterns of moral evaluations that rely on these very processes, as compared to the neurotypical population. In addition to these deficits, prior work on neuroanatomical and neurofunctional correlates of MS reveals that the disease presents with: (*a*) frontotemporal changes such as atrophy (Bonavita, Tedeschi, & Gallo, 2013; Calabrese et al., 2010); (*b*) affected structural (Fox, 2008) and (*c*) functional (Sacco, Bonavita, Esposito, Tedeschi, & Gallo, 2013) frontotemporal connections; and (*d*) aberrant amygdala-frontotemporal activation at the resting state (Nigro et al., 2015), and also during the processing of salient affective stimuli (Passamonti et al., 2009; Sacco et al., 2013). Given that research with neurological populations with involvement of frontotemporal regions have revealed deficits in intent-based moral judgments (Baez, Couto, et al., 2014; Baez, Manes, et al., 2016; Baez, Kanske, et al., 2014; Baez, Morales, et al., 2016; Ciaramelli et al., 2012; Young, Bechara, et al., 2010), we expected a similar disruption in MS. Thus far, only one study has investigated moral cognition in the MS population (Gleichgerrcht et al., 2015). MS patients and healthy controls (HC) were presented with moral dilemmas asking about the permissibility of sacrificing the welfare of the few in favor of the aggregate welfare; MS patients judged the act of sacrificing the few for the greater good (utilitarian response) as less morally permissible due to increased emotional reactivity. This is a surprising result given that MS patients showed elevated levels of alexithymia and reduced EC, both of which are associated with increased utilitarian inclinations (Gleichgerrcht & Young, 2013; Koven, 2011; Patil, Melsbach, Hennig-Fast, & Silani, 2016; Patil & Silani, 2014a). Additionally, MS patients also showed elevated egocentric moral attitudes (compared to HC) in that they reported that others would respond the same way as they did.

In the current exploratory study, we extended this work with a task that featured relatively ordinary and familiar scenario settings (as compared to more contrived moral dilemma contexts). Furthermore, the task featured not only intentional harms, but also unintentional or accidental harms, and behaviors that were performed without the goal of maximizing aggregate welfare. This task may therefore probe the subtle effects of social cognition deficits on moral cognition in MS patients. In the light of the prevalent ToM deficits, on the one hand, we expected MS patients to provide lenient judgments of attempted harms or harsh judgments of accidental harms. On the other hand, empathic deficits were expected to lead to lenient judgments of accidental harms. In light of the complicated pattern of social cognitive deficits in MS with divergent associated predictions, we did not have more specific predictions and relied on exploratory data analysis.

2. Materials and methods

2.1. Participants

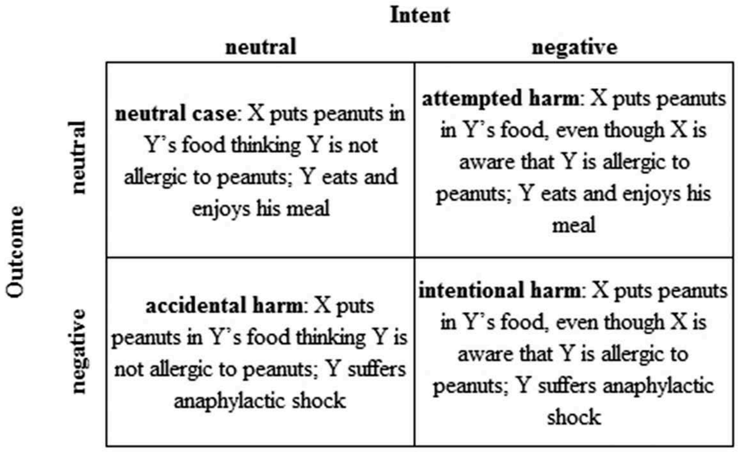
Thirty-eight consecutive patients (86.8% female) who fulfilled the McDonald criteria (Polman et al., 2011) for Relapsing-Remitting Multiple Sclerosis (RRMS) were recruited for the present study. A large portion of these patients were previously recruited for a study on moral judgment (Gleichgerrcht et al., 2015). Patients reported no history of alcohol/drug abuse, major psychiatric disorder, or traumatic brain injury. They were assessed at least 90 days after the most recent relapse episode and had been off steroid treatment for at least three months. Patients who were having a relapse in their MS were not included in the study. All the patients were receiving disease-modifying therapies at the time of assessment and scored above the proposed cutoff score for the Mini-Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975; score >24). Their mean Expanded Disability Status Scale (EDSS) score (Kurtzke, 1983) was 1.66 (SD = 1.6, range = 0–6, median = 1.25), indicating mild MS. Mean disease duration (in number of years) was 10.60 (SD = 8.7, range = 1.38–39.3, median = 9.01). Mean number of relapses was 3.4 (SD = 1.92, range = 2–12, median = 3), and mean Multiple Sclerosis Severity Score (MSSS) (Roxburgh et al., 2005) was 2.35 (SD = 2.4, range = 0.01–8.64, median = 1.29) points.

Thirty-eight age-, gender-, and level of education-matched volunteers were also included in the HC group after having undergone screening to ensure absence of history of drug abuse, neurological or neuropsychiatric disorders. Neither MS patients nor HC were financially compensated and voluntarily participated in the study. All participants signed an informed consent form before participating in this study, which was approved by the Ethics Committee at the Institute of Cognitive Neurology (INECO, Buenos Aires, Argentina).

2.2. Empathy and alexithymia measurement

The Interpersonal Reactivity Inventory (IRI) (Davis, 1983; Spanish-validated version: Pérez-Albéniz, De Paúl, Etxeberría, Montes, & Torres, 2003) was used to assess specific aspects [fantasizing, EC, perspective-taking (PT), and PD] of dispositional empathy. Inspired by recent theoretical and empirical work involving empathy and moral cognition (Decety & Cowell, 2014; Decety & Yoder, 2015), we used EC, PT, and PD subscales to denote motivational, cognitive, and affective components of empathy, respectively. Moreover, based on recent psychometric assessments of the IRI questionnaire (Baldner & McGinley, 2014), we decided a priori not to explore the fantasy subscale beyond descriptive statistics, as it does not map well onto the current theorizations of empathy.

To assess the levels of trait alexithymia, we used the validated Spanish version of the Toronto Alexithymia Scale-20 (TAS-20) questionnaire (Bagby, Taylor, & Parker, 1994; Spanish version: Martínez Sánchez, 1996) consisting of three subscales: Difficulty Describing



		Intent	
		neutral	negative
Outcome	neutral	**neutral case**: X puts peanuts in Y's food thinking Y is not allergic to peanuts; Y eats and enjoys his meal	**attempted harm**: X puts peanuts in Y's food, even though X is aware that Y is allergic to peanuts; Y eats and enjoys his meal
	negative	**accidental harm**: X puts peanuts in Y's food thinking Y is not allergic to peanuts; Y suffers anaphylactic shock	**intentional harm**: X puts peanuts in Y's food, even though X is aware that Y is allergic to peanuts; Y suffers anaphylactic shock

Figure 1. Four types of possible harms (conditions) from a 2 (belief: neutral, harmful) × 2 (outcome: neutral, harmful) design. One example is also shown in the form of an abbreviated story. Note that in the current study each scenario appeared only in one condition. See Supplementary Text S2 to access the full list of stories and the conditions in which they appeared.

Feelings (DDF, five items, e.g., "I am often puzzled by sensations in my body."), Difficulty Identifying Feelings (DIF, seven items, e.g., "I often don't know why I am angry."), and Externally Oriented Thinking (EOT, eight items, e.g., "I prefer to analyze problems rather than just describe them"). Out of these three constructs, DDF and DIF together represent the affective aspects of alexithymia (Zackheim, 2007), while EOT is construed as a cognitive/attentional component of alexithymia as it is less affective in its scope (Moriguchi & Komaki, 2013). Following the criteria proposed by the original authors (Bagby et al., 1994), individual scores were used to classify participants as either non-alexithymic (score ≤51), borderline alexithymic (scores of 52–60), or as alexithymic (scores ≥61).

For more details about these questionnaires and their internal reliability analysis, see Supplementary Text S1.

2.3. Moral judgment task

Experimental stimuli consisted of 24 unique stories, equally divided among four conditions resulting from a 2 (belief: neutral, negative) × 2 (outcome: neutral, negative) within-subject design such that agents in the scenario produced either a neutral outcome or a harmful outcome while acting with the belief that they were causing either a neutral outcome or a harmful outcome (see Figure 1). The magnitude of harm severity varied freely across scenarios from mild, to severe, to fatal injuries. We note that the stories were not balanced across conditions, i.e., a given story appeared only in one condition (e.g., accidental), but never in any other condition. Stories were matched for word length across conditions. All scenarios were adapted in Spanish from Young, Camprodon, et al. (2010) (see Supplementary Text S2 to see the scenario-by-condition breakdown and labels for scenarios from original material).

Each scenario was presented on screen for as long as the patient needed in order to make sure that the subjective feeling of time pressure did not affect responses and also to alleviate working memory demands for patients. Each story consisted of four cumulative segments: (*i*) *background*: this segment provided the setting in which the story took place; (*ii*) *foreshadow*: this segment foreshadowed whether the outcome would be neutral or harmful; (*iii*) *belief*: this segment provided information about whether the agent was acting with a neutral or harmful belief; (*iv*) *consequence*: this final segment revealed the outcome of the agent's action.

We provide below an example of one scenario ("Vitamin") that appeared in the accidental harm condition:

Juan is instructed by a doctor to give his senile wife pills for her heart disease. The doctor says that she must not intake vitamin K within an hour to take the pills safely. One day, his wife tries a new kind of fruit. The new kind of fruit is high in vitamin K, so it is deadly for Juan's wife to take the pills right away. Juan does his research and believes that the new kind of fruit does not have vitamin K and that it is safe to give her the pills. Juan gives his wife the pills right away. His wife dies of heart failure.

The order of presentation of scenarios was randomized across subjects. After reading each scenario, participants provided three types of moral judgments, which always appeared in the same order:

(i) *appropriateness*: “How appropriate was it for [the agent] to do [nature of the action]?” (7-point Likert scale: 1 = completely inappropriate; 7 = completely appropriate);

(ii) *punishment*: “How severely should [the agent] be punished for [nature of the action]?” (7-point Likert scale: 1 = no punishment, 7 = severe punishment);

(iii) *egocentrism*: “Out of 100 people answering to this scenario, how many do you think would answer like you?” (continuous scale: from 0 to 100).

Appropriateness ratings were later reverse-scored to be positively associated with the punishment judgments, and higher scores thus reflect perceived *inappropriateness* of the agent’s behavior. Participants had up to 30 s to respond to each question.

2.4. Statistical analysis

Data was analyzed using SPSS 22 (for classical inference) and JASP 0.7.4 (for Bayesian inference). Effect size measures are reported (Lakens, 2013, 2015a) along with Bayes factors (cf., recommendations by Lakens, 2015b). We follow guidelines provided in Nimon (2012) to ensure that our data met the statistical assumptions associated with the general linear model-based statistical tests that we employed. Correlation analysis was carried out using Spearman’s *rho* since it is more robust to univariate outliers than Pearson’s *r* (Pernet, Wilcox, & Rousselet, 2012). Welch’s *t*-test was used as a default for between-group comparisons instead of Student’s *t*-test because it accounts for unequal variances between groups (cf., Lakens, 2015c). As recommended (Weissgerber, Milic, Winham, & Garovic, 2015), in addition to bar graphs in the main text, we have also provided univariate scatter-plots for main dependent variables in the supplementary material.

3. Results

3.1. Demographic information

The two groups were well-controlled for demographic variables as they did not differ in terms of their age (HC = 39.3 (8.1), MS = 42.3 (11.3); $t(66.69) = 1.309$, $p = 0.195$, $BF_{10} = 0.497$), number of years of formal education (HC = 15.7 (1.8), MS = 15.4 (2.8); $t(64.16) = -0.680$, $p = 0.499$, $BF_{10} = 0.290$), or gender composition (proportion of females; HC = 81.60%, MS = 86.80%; $\chi^2(1) = 0.396$, $p = 0.529$, $BF_{10} = 0.367$).

3.2. Group differences in alexithymia and empathy

As expected from prior work (Gleichgerrcht et al., 2015), we found that patients with MS showed elevated levels of all aspects of alexithymia (Table 1). There was a significant difference between individuals in HC and MS groups ($\chi^2(2) = 24.346$, $p < 0.001$, $BF_{10} = 41975$) in the composition of alexithymic (HC = 1/38 vs. MS = 12/38), borderline alexithymic (HC = 1/38 vs. MS = 10/38), and nonclinically alexithymic (HC = 36/38 vs. MS = 16/38). Moreover, the MS population showed reduced scores on the motivational component of empathy (EC), increased levels of affective sharing (PD), and did not differ on the cognitive component of empathy (PT) relative to the HC group (see Table 1).

We ensured that these differences were not due to the poor internal reliability of measures in the MS population by checking Cronbach’s alpha values for each subscale and for each group (see Supplementary Text S1).

3.3. Descriptive statistics for moral judgments

Before averaging judgments across scenarios in each condition at the subject level, we ascertained that the items showed good internal reliability (see Supplementary Text S3). Average ratings for all 24 scenarios for each group are provided in Supplementary Text S4.

Table 1. Group differences in levels of alexithymia and empathy. Values in parentheses denote standard deviation.

Variable	MS (*n* = 38)	HC (*n* = 38)	*t*	*df*	*p*	Cohen’s *d*	BF10
DDF	13.95 (4.17)	10.89 (3.73)	3.364	73.13	0.001	0.772	25.77
DIF	15.97 (7.57)	11.45 (4.67)	3.137	61.62	0.003	0.72	14.34
EOT	27.45 (4.83)	15.82 (4.23)	11.175	72.75	<.001	2.564	2.389e +14
TAS-20	57.37 (13.40)	38.16 (9.71)	7.156	67.46	<.001	1.642	1.563e +7
PT	3.602 (0.642)	3.425 (0.645)	1.196	74	0.235	0.274	0.44
F	2.891 (0.795)	3.41 (0.846)	−2.754	73.72	0.007	−0.632	5.762
EC	3.342 (0.67)	3.816 (0.624)	−3.188	73.63	0.002	−0.731	16.287
PD	3.066 (0.724)	2.544 (0.812)	2.957	73.07	0.004	0.678	9.233

Table 2. Mean (SD) for appropriateness, punishment, and egocentric moral judgments for each group.

Type of judgment	Scenario	Group	Mean (SD)
Appropriateness	Neutral	HC	2.259 (1.029)
		MS	2.618 (0.918)
	Accidental	HC	2.197 (0.856)
		MS	2.693 (1.125)
	Attempted	HC	6.224 (0.735)
		MS	6.289 (1.037)
	Intentionl	HC	6.509 (0.517)
		MS	6.645 (0.653)
Punishment	Neutral	HC	1.544 (0.629)
		MS	2.105 (0.974)
	Accidental	HC	1.684 (0.75)
		MS	2.259 (0.992)
	Attempted	HC	5.004 (1.271)
		MS	5.548 (1.319)
	Intentionl	HC	5.509 (1.167)
		MS	6.18 (1.055)
Egocentric	Neutral	HC	79.123 (16.184)
		MS	80.246 (13.561)
	Accidental	HC	71.382 (15.582)
		MS	79.803 (13.961)
	Attempted	HC	67.75 (14.766)
		MS	82.522 (14.209)
	Intentionl	HC	76 (13.486)
		MS	87.93 (13.286)

Descriptive statistics for appropriateness, punishment, and relative moral judgments averaged at the group-level are reported in Table 2.

3.4. Group differences in moral judgments about appropriateness

To assess whether the two groups relied to a different degree on belief and outcome information while evaluating the moral appropriateness of the actions, we conducted a 2(belief) × 2(outcome) × 2(group) mixed ANOVA. This analysis revealed a main effect of belief (F(1,74) = 1120.59, $p < 0.001$, $p\eta^2 = 0.938$, $\omega^2 = 0.935$) and outcome (F(1,74) = 4.838, $p = 0.031$, $p\eta^2 = 0.061$, $\omega^2 = 0.048$) but no group-by-belief (F(1,74) = 2.029, $p = 0.173$, $p\eta^2 = 0.025$, $\omega^2 = 0.001$) or group-by-outcome (F(1,74) = 0.481, $p = 0.490$, $p\eta^2 = 0.006$, $\omega^2 \sim 0$) interactions. In other words, the behavior of agents who were acting with a harmful intent was judged to be more inappropriate as compared to agents who were acting with a neutral intent (see Figure 2). Similarly, agents who produced harmful outcome were condemned more harshly as compared to agents who did not. More importantly, MS patients did not differ from the control population in the degree to which they relied on outcome and intent information while deciding on how appropriate the agent's behavior was (see Figure 2).

These main effects were qualified by an interaction effect between belief and outcome (F(1,74) = 5.141, $p = 0.026$, $p\eta^2 = 0.065$, $\omega^2 = 0.052$), but not a belief-by-outcome-by-group interaction (F(1,74) = 0.057, $p = 0.813$, $p\eta^2 = 0.001$, $\omega^2 \sim 0$). These results are indicative of the fact that moral judgments about negative and neutral outcomes depended on whether these outcomes were produced with negative or neutral belief, and this dependence did not differ across groups (Figure 1). In particular, attempted harms were found to be less appropriate than the accidental harms based on the harmful intent ($ps < 0.001$). This oft-reported pattern of judgments from the healthy population was also preserved in MS patients. One surprising result was that the behavior of accidental harm-doers was judged to be just as appropriate as the agents involved in neutral condition, but, more importantly, this was the case for both MS and HC groups (see Table 3).

There was also a main effect of group (F(1,74) = 4.243, $p = 0.043$, $p\eta^2 = 0.054$, $\omega^2 = 0.041$), such that MS patients made harsher judgments than the HC (for additional details and scatter-plots, see Supplementary Text S5). Patients with MS generated higher average wrongness ratings than the HC group for 16 (out of 24) scenarios (see Supplementary Text S4). To investigate the source of this main effect, we carried out exploratory analyses in which we repeated the mixed-effects ANOVA separately for each independent variable that differed between two groups, viz. EC, PD, DDF, DIF, and EOT (see Table 1). Only in the ANCOVA with EOT as the covariate did the main effect of group vanish ($\omega^2 = 0.036$, $p > 0.05$). In other words, although MS patients were not impaired in assessing the appropriateness of specific types of harmful acts (intentional, attempted, etc.), they were nonetheless harsher *in general* in judging third-party moral violations. Furthermore, once intergroup variance associated with externally oriented cognition (EOT) was removed, the difference between MS patients and HC was no longer significant.

3.5. Group differences in moral judgments about punishment

The pattern of results obtained for appropriateness judgments was also observed for punitive judgments (see Figure 3). A 2(belief) × 2(outcome) × 2(group) mixed ANOVA revealed a main effect of belief (F(1,74) = 714.63, $p < 0.001$, $p\eta^2 = 0.906$, $\omega^2 = 0.906$), outcome (F(1,74) = 22.495, $p < 0.001$, $p\eta^2 = 0.233$, $\omega^2 = 0.233$), and also an interaction effect between belief and outcome (F(1,74) = 11.196, $p = 0.001$, $p\eta^2 = 0.131$, $\omega^2 = 0.131$). More importantly, none of these effects interacted with group factor (belief-by-group: F(1,74) = 0.021, $p = 0.886$, $p\eta^2 \sim 0$, $\omega^2 \sim 0$; outcome-by-group: F(1,74) = 0.217, $p = 0.643$, $p\eta^2 = 0.003$,

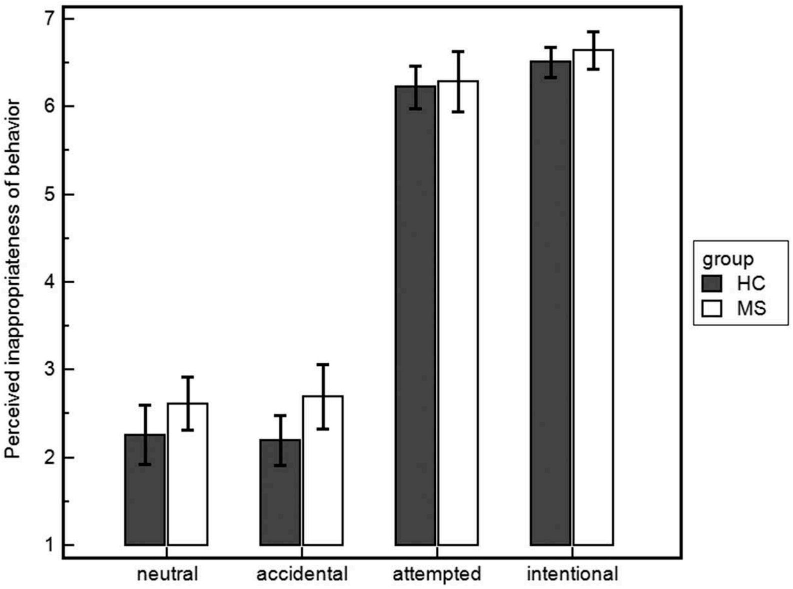


Figure 2. Moral judgments about appropriateness of behavior of moral agents by MS patients and healthy controls on a 7-point Likert scale (1: *completely appropriate*, 7: *completely inappropriate*) for different types of harms: neutral case (neutral belief, neutral outcome), accidental harm (neutral belief, negative outcome), attempted harm (negative belief, neutral outcome), and intentional harm (negative belief, negative outcome). Error bars represent 95% confidence intervals.

Table 3. Bonferroni-corrected post hoc comparisons between neutral/negative belief and neutral/negative outcome cases for MS and HC groups.

		Appropriateness		Punishment	
Comparison	Group	Mean difference	*p*-value	Mean difference	*p*-value
Accidental-neutral	HC	−0.061	0.681	0.14	0.436
	MS	0.075	0.667	0.154	0.484
Intentional-attempted	HC	0.285	0.042	0.504	<0.001
	MS	0.355	0.020	0.632	0.002
Attempted-neutral	HC	3.965	<0.001	3.461	<0.001
	MS	3.671	<0.001	3.443	<0.001
Intentional-accidental	HC	4.311	<0.001	3.825	<0.001
	MS	3.952	<0.001	3.921	<0.001

ω^2 = 0.003; belief-by-outcome-group: $F(1,74)$ = 0.205, p = 0.652, $p\eta^2$ = 0.003, ω^2 = 0.003).

Put differently, while deciding on punishment for agents involved in moral situations, MS and HC groups relied to an equal extent on both belief and outcome information (for comparisons between different cases, see Table 3; for scatter-plots, see Supplementary Text S5).

Importantly, there was once again a main effect of group ($F(1,74)$ = 11.92, p < 0.001, $p\eta^2$ = 0.139, ω^2 = 0.126) such that MS delivered harsher moral judgments than controls (see Figure 3; for scatter-plots and additional details, see Supplementary Text S5). Strikingly, patients with MS generated higher average punishment ratings than the HC group for 22 (out of 24) scenarios (see Supplementary Text S4). In other words, not only did MS patients find the behavior of agents to be less appropriate than the HC group (as shown by the harsher appropriateness judgments across the board), but they were also more punitive than the control participants.

In exploratory analyses, we repeated this mixed ANOVA separately for each independent variable that differed between two groups, viz. EC, PD, DDF, DIF, and EOT (see Table 1), to probe the source of this effect. All group differences remained significant in ANCOVAs with EC, PD, DDF, and DIF as covariates. Once again only in ANCOVA with EOT as the covariate, although the main effect of group was still significant (p = 0.014), none of the planned group comparisons were significant (see Supplementary Text S5) and the effect size for the main effect of group was reduced by almost half (from ω^2 = 0.126 to ω^2 = 0.063). In other words, differences in EOT scores significantly accounted for elevated punitive judgments in MS patients than the controls.

3.6. *Distinction between appropriateness and punishment judgments across groups*

Prior work has shown that punishment judgments rely to a greater degree on outcome information as compared to appropriateness judgments (Cushman, 2008). In order to investigate whether this distinction was observed in the current study and whether it was preserved in the MS population, we conducted a 2(belief) × 2(outcome) × 2 (type of judgment: appropriateness, punishment) × 2 (group) mixed ANOVA. Indeed, this analysis revealed an

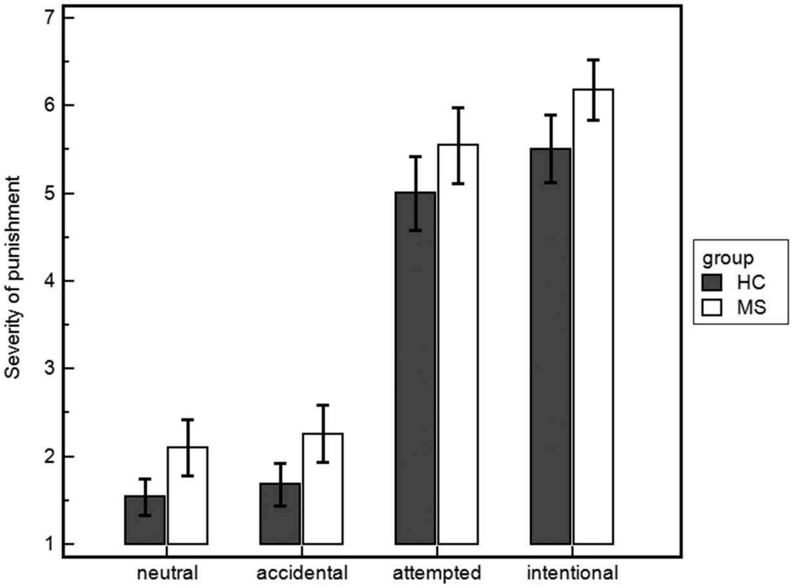


Figure 3. Punitive judgments endorsed by MS patients and healthy controls on a 7-point Likert scale (1: *no punishment*, 7: *severe punishment*) for different types of harms: neutral case (neutral belief, neutral outcome), accidental harm (neutral belief, negative outcome), attempted harm (negative belief, neutral outcome), and intentional harm (negative belief, negative outcome). Error bars represent 95% confidence intervals.

outcome-by-judgment interaction (F(1,74) = 4.623, p = 0.035, $p\eta^2$ = 0.059, ω^2 = 0.083), but no outcome-by-judgment-by-group interaction (F(1,74) = 0.056, p = 0.813, $p\eta^2$ = 0.001, ω^2 ~ 0). To explore this effect further, we separately carried out a 2(outcome) × 2 (judgment) × 2(group) mixed ANOVA for neutral (accidental harm vs. neutral cases) and negative (intentional vs. attempted harm cases) belief, but no outcome-by-judgment-by-group interaction was found either for neutral (F(1,74) = 0.492, p = 0.485, $p\eta^2$ = 0.007, ω^2 ~ 0) or for negative (F(1,74) = 0.077, p = 0.782, $p\eta^2$ = 0.001, ω^2 ~ 0) belief.

To summarize, consistent with prior work (Cushman, 2008), participants assigned greater weight to information about the nature of outcomes (harmful or neutral) more heavily when deciding on punishment (ω^2 = 0.233) as compared to judging appropriateness (ω^2 = 0.048), irrespective of whether the intent was neutral or harmful. Interestingly, this distinction between moral judgments was preserved in the MS patients. The same conclusion can be drawn from the percent of variance[1] explained by each factor for each type of judgment for the HC and MS groups (see Table 4).

Table 4. Percentage of total variability explained by belief and outcome factors and their interactions for each type of moral judgment and for each group.

	Appropriateness		Punishment	
Factor	HC	MS	HC	MS
Belief	91.68	87.03	87.07	84.00
Outcome	0.07	0.28	0.68	0.95
Belief-by-outcome	0.16	0.12	0.22	0.35
Error	8.09	12.58	12.03	14.69

This is a striking result as the MS patients scored lower than controls on EC and more on PD, and thus one would have expected them to be, respectively, either less or more reliant on outcomes while endorsing punishment judgments (vis-à-vis appropriateness judgments) as compared to controls. We note, however, that MS patients did not exhibit any reduction in the cognitive component of empathy (PT). This led us to suspect that this component of PT may play a crucial role in mediating the influence of harmful outcome on punishment judgments to a greater degree than appropriateness judgments. In other words, people may deliver harsher punishment versus appropriateness judgments

[1] Note that belief accounted for a surprising 84–91% of total variability, while outcome accounted for less than 1% of variance for both type of judgments. This contrasts with prior findings, which show that outcome explains up to 3% (for wrongness) or 20% (for punishment) of the total variation in moral judgment (e.g., Cushman, 2008). This pattern can also be discerned for appropriateness and punishment judgments where the agents who accidentally harmed someone were not judged more severely than the neutral cases by both groups. At present, we cannot distinguish whether this departure from past findings is due to the cultural settings in which the experiment was conducted (Argentina vs. US) or artefact of limited sample size (n = 38 vs. n > 1000) or differences in design (within-subject vs. between-subject).

for accidental harm-doers compared to neutral cases because of an empathic aversion stemming from victim PT (and not EC or affective sharing) that motivates them to punish. This hypothesis predicts that, once the inter-individual variation in PT is accounted for, we should not observe interaction between the nature of outcome and type of moral judgment. Accordingly, we carried out a 2(belief) × 2(outcome) × 2(type of judgment) repeated-measures ANOVA, yielding the expected interaction effect between outcome and type of judgment ($F(1,75) = 7.958$, $p = 0.006$, $p\eta^2 = 0.096$), but this interaction was no longer significant when PT was added as a covariate to the model ($F(1,74) = 0.130$, $p = 0.720$, $p\eta^2 = 0.002$).

Because this was an unexpected and novel result, we replicated this effect in a larger sample ($n = 113$) consisting of only healthy adults and with a larger battery of scenarios ($n = 36$). Replication of this effect was crucial for two additional reasons: (*i*) the moral judgments observed in the current sample were unusually dependent on intent information (as shown by equal punishment for accidental and neutral cases; cf. Table 4), and (*ii*) the scenarios were not counterbalanced across conditions. This new study addressed these concerns and successfully replicated the observed effects (see Appendix). Thus, we conclude that cognitive empathy (or PT) plays a key role in the relatively greater influence of outcome information on punishment versus appropriateness judgments.

3.7. Group differences in egocentric moral judgments

In order to assess group differences in egocentric moral judgments (also see Table 2), we ran a 2(belief)-by-2(outcome)-by-2(group) mixed ANOVA, which revealed a main effect of group such that MS patients reported that others would agree with their judgments to a greater degree than controls, irrespective of the type of scenario ($F(1,74) = 10.590$, $p = 0.002$, $p\eta^2 = 0.054$, $\omega^2 = 0.112$). Thus, MS patients were more confident than the HC that others would agree with their judgments (see Figure 4; for scatter-plots and additional details, see Supplementary Text S5). Indeed, patients with MS generated higher egocentric ratings than the HC group for 20 (out of 24) scenarios (see Supplementary Text S4).

As with the punishment judgments, we carried out mixed ANCOVA separately for each covariate of interest (EC, PD, DDF, DIF, and EOT) to explore the possible source of this effect. Once again, this analysis showed that the main effect of group was no longer significant ($\omega^2 \sim 0$, $p > 0.05$) only when EOT was added as a covariate to the model.

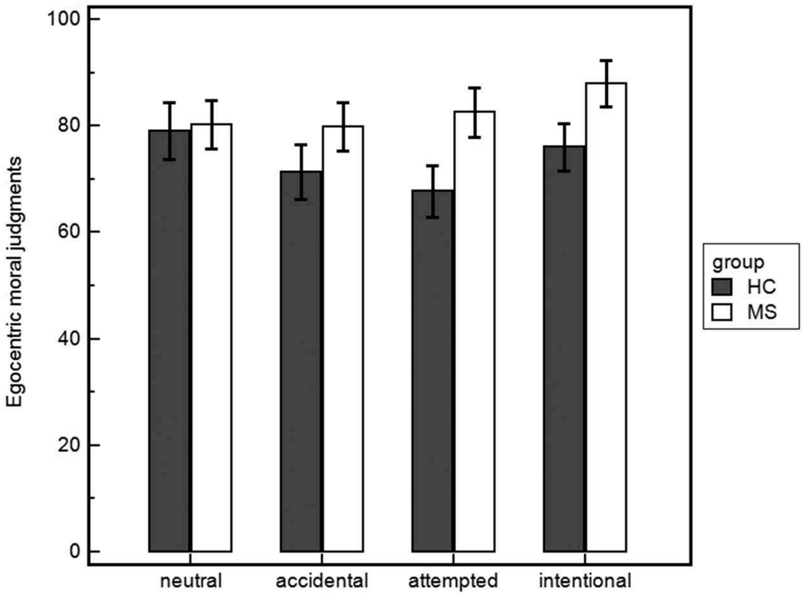


Figure 4. Egocentric moral judgments by MS patients and healthy controls to assess the expected degree of congruency between one's own response and others' responses ("Out of 100 people answering to this scenario, how many do you think would answer like you?"), for different types of harms: neutral case (neutral belief, neutral outcome), accidental harm (neutral belief, negative outcome), attempted harm (negative belief, neutral outcome), and intentional harm (negative belief, negative outcome). The responses were collected using a continuous scale whereby participants could choose any number from 0 to 100. Error bars represent 95% confidence intervals.

3.8. Correlation analysis: relation between clinical variables, alexithymia, empathy, and moral judgments

Given that the sample size was limited for each group (n = 38), we provide preliminary analyses of the relationships among interindividual differences in personality traits and clinical or dependent variables of interest in Supplementary Text S6.

3.9. Correlation analysis: relation between appropriateness and punishment judgments

As would be expected, there was a high positive correlation between appropriateness and punishment judgments for a given condition, and these associations were not different between controls and MS patients, as shown by Fisher's *Z*-test (see Supplementary Text S7 for full analysis).

4. General discussion

In the current study, we investigated how the known impairments in social cognition processes in MS patients affect their moral judgments in contexts where information about beliefs and outcomes conflict. The findings revealed a complex pattern of moral judgments that deviate from HC. We discuss each result individually in the following subsections.

4.1. Preserved moral judgments in MS for specific acts

Past evidence has shown that both condemning agents who attempt but fail to cause harm and exculpating agents who accidentally cause harm depend on recruitment of the ToM network (e.g., Young et al., 2007, Young, Camprodon, et al., 2010). This research has also underscored that mental state reasoning is especially critical for the latter process—exculpation in the case of accidents. In the case of a prepotent tendency to condemn the agent due to her causal responsibility for the negative outcome (e.g., harm), a robust representation of the agent's innocent mental state is needed to override this response (Moran et al., 2011; Young et al., 2007; Young & Saxe, 2009). This basic model also finds support from studies of clinical populations with ToM impairments, like autism; individuals with autism properly assess the wrongness of attempted harms but condemn accidental harms more (Buon, Dupoux, et al., 2013; Moran et al., 2011; Salvano-Pardieu et al., 2015). Thus, we expected a similar pattern among MS patients: elevated condemnation for accidental harm, but no difference for attempted harm.

In the current study, however, we did not observe any group difference for the accidental harm cases. One possible explanation for this null result is that MS patients exhibit deficits not only in ToM but also in EC. Past research shows that more empathic individuals (especially those scoring high on EC) find accidental harms to be more wrong/unacceptable (Patil & Silani, 2014b; Trémolière & Djeriouat, 2016; Walter et al., 2012), whereas clinical populations with reduced empathy are more lenient in their moral judgments of accidental harms (alexithymia: Patil & Silani, 2014b; psychopathy: Young et al., 2012). Since there is likely to be no strong negative affect stemming from empathic aversion that would otherwise lead to increased condemnation of accidental harm cases in MS patients, following the two-process framework, no robust mental state representation is needed to forgive the agent. Specifically, the ToM and empathy deficits in MS patients likely exert mutually opposite influence on the final judgment that cancel each other out and leave their final appropriateness judgments relatively intact (reduced ToM ➔ accidental harms more wrong; reduced EC ➔ accidental harms less wrong; final output ➔ preserved moral assessment of accidental cases). In the case of MS, reduced empathy alongside ToM deficits therefore contribute to an apparently typical pattern of moral judgment. Note that of all the personality traits and clinical populations studied so far on this task (autism, alexithymia, psychopathy), MS remains the only one that has deficits in ToM *and* empathy and not ToM *or* empathy. We also note that, although empathic reasoning and mentalizing are two independent processes that contribute their unique inputs (Cushman, 2008), these inputs need to be properly integrated for the final moral judgments (Buon, Jacob, et al., 2013). Impairment in any one process (e.g., mentalizing) will leave the input from other process intact (e.g., although mentalizing is impaired in autism, the input from empathizing is preserved), and this will affect the final output (increased condemnation for accidents). We propose that, in case of MS (as compared to HC), since both processes are impaired, the final output is comparable to that in HC.

4.2. Role of EOT in harsher global moral judgments in MS

Although MS patients did not differ from controls in terms of their reliance on belief and outcome information while endorsing moral judgments, they nevertheless found the third-party behaviors (regardless of the type of scenario) to be less appropriate and more deserving of punishment and this effect was driven by EOT. Note that neither ToM nor empathy deficits predict this pattern for reasons mentioned in the preceding section (Section 4.1).

One possible explanation is that this harsher moral attitude toward others' behavior stems from the affective disorders (depression, anxiety, and fatigue) that are prevalent and persistent in MS patients (e.g., Bodini et al., 2008; Chahraoui et al., 2014; Gay et al., 2010; for a review see Sá, 2008). MS is a chronic disease without a cure that usually strikes in the most formative years of one's life (20–40 years). Additionally, the long-term course of MS highly varies from patient-to-patient; some patients can go without an attack for years, while others can go from being perfectly healthy to having a rapid physical and cognitive decline in a short period of time. This unpredictable and ambiguous nature of progression in MS is considered to be the fountainhead of affective disorders in MS (Chahraoui et al., 2014). When patients are in remission phase, they tend to be constantly in a state of fear and anxiety due to the highly unpredictable nature of their disease course, uncertain of when the next attack (relapse) can flare up; in addition, during the course of a relapse, they worry about possible imminent disability. Extant research (Chahraoui et al., 2014) suggests that one possible defensive strategy that patients use to cope with the traumatic experience from past attacks (Sá, 2008) is to develop a trait-like pattern of orienting their thoughts on the external events rather than introspecting on their inner experiences and feelings (i.e., EOT). Although alexithymia, in general, has been argued to be an inhibition mechanism that acts as an anti-stress fortification in both healthy and clinical populations, with different dimensions of alexithymia playing a role in specific situations or in response to specific stimuli (Davydov, Stewart, Ritchie, & Chaudieu, 2010), EOT seems to be the locus of this mechanism. Studies using multiple approaches have revealed evidence for the role of EOT as an adaptive strategy that leads to the avoidance of self-reflection and thereby blunts the influence of negative arousal states to protect individuals in the long term from stress (for a more detailed discussion, see Demers & Koven, 2015). For example, EOT has been found to distract attention from self-oriented ruminations toward external environment, thereby disrupting repetitive negative thoughts that exacerbate sad or depressed mood while protecting the individual from negative arousal states (Luminet, Rimé, Bagby, & Taylor, 2004). EOT is also associated with an inability to adopt an experiential mode of thinking (Di Schiena, Luminet, & Philippot, 2011). Psychophysiological studies demonstrate the moderating role of EOT in decoupling subjective feeling states from physiological arousal states (Davydov, Luminet, & Zech, 2013), and behavioral studies demonstrate that DDF and DIF, but not EOT, predict levels of depression and anxiety in HC (Hendryx, Haviland, & Shaw, 1991). Additional evidence supporting the role of EOT as a cognitive, adaptive strategy comes from clinical populations with affective disorders in which DDF and DIF—but not EOT—account for variance in depression and anxiety, e.g., posttraumatic stress disorder (Monson, Price, Rodriguez, Ripley, & Warner, 2004), childhood abuse and neglect (Güleç et al., 2013), borderline personality disorder (Evren, Cınar, & Evren, 2012), depressive and anxiety disorders (Marchesi, Brusamonti, & Maggini, 2000; Saarijärvi, Salminen, & Toikka, 2001), etc. Similarly, implication of EOT as an adaptive strategy in MS population explains why DIF and DDF factors of alexithymia, but not EOT, are found to predict depression, anxiety, and fatigue in MS patients (Bodini et al., 2008; Chahraoui et al., 2014).

Given this reduced ability to effectively process their own emotional states and a cognitive style that is externally oriented, MS patients are expected to have poor emotion differentiation skills (Lindquist & Barrett, 2008), which would lead them to represent their affective states only in global terms. Therefore, they would not be able to disentangle affect *integral* to the scenario (in response to harmful intent and/or outcome) from *incidental* affect (cf., Cameron, Payne, & Doris, 2013). Both integral and incidental affect have been well-known to influence judgment and decision making (for a review, see Västfjäll et al., 2016). Thus, the prevalent negative mood states in MS can be misattributed by patients to the experimental stimulus and this negative emotional arousal, in turn, would amp up the severity of moral condemnation (Cheng, Ottati, & Price, 2013). Västfjäll and colleagues have argued that the incidental affect can have an especially greater bearing on judgments when the incidental affect is salient but the source awareness is weak. This is indeed the case for MS patients who experience salient negative affective states, but the EOT strategy reduces their awareness of source of this affect.

The current data support this hypothesis given that the main effect for the appropriateness judgments was not significant and significantly reduced for punishment judgment only after controlling for the group differences in the EOT component of alexithymia (Bodini et al., 2008; Chahraoui et al., 2014). Thus, the strategy that MS patients employ to protect themselves from distressing affective states seems to also prevent them from insulating their moral judgments[2] from the prevalent, incidental negative affect that leads them to make harsher judgments as compared to HC (see Figure 5 for schematic illustration). In other words, the harsher moral attitudes in MS seem to be not due to MS affectation per se, but due to comorbid alexithymia, especially its subcomponent EOT.

Note that this interpretation still does not explain why the same effect was relatively weak for wrongness judgments, i.e., why the incidental negative arousal affected punishment judgments to a greater extent (effect size: ω^2 = 0.126) than wrongness judgments (effect size: ω^2 = 0.041). We propose that this is due to the fact that blame/punishment judgments in general rely to a greater degree on affective processes than wrongness/appropriateness judgments, as shown by both neuroimaging (Buckholtz et al., 2008, 2015; Treadway et al., 2014) and behavioral (Cushman, 2008; also see Malle, Guglielmo, & Monroe, 2014) studies.

4.3. *Increased egocentric moral attitudes in MS*

Interestingly, MS patients were more likely to report that other people would evaluate third-party moral violations the same way as they do than the HC. In other words, MS patients displayed more self-oriented moral attitudes and could not foresee that others may hold dissenting moral opinions about the behavior of agents in these scenarios. Interestingly, this effect was

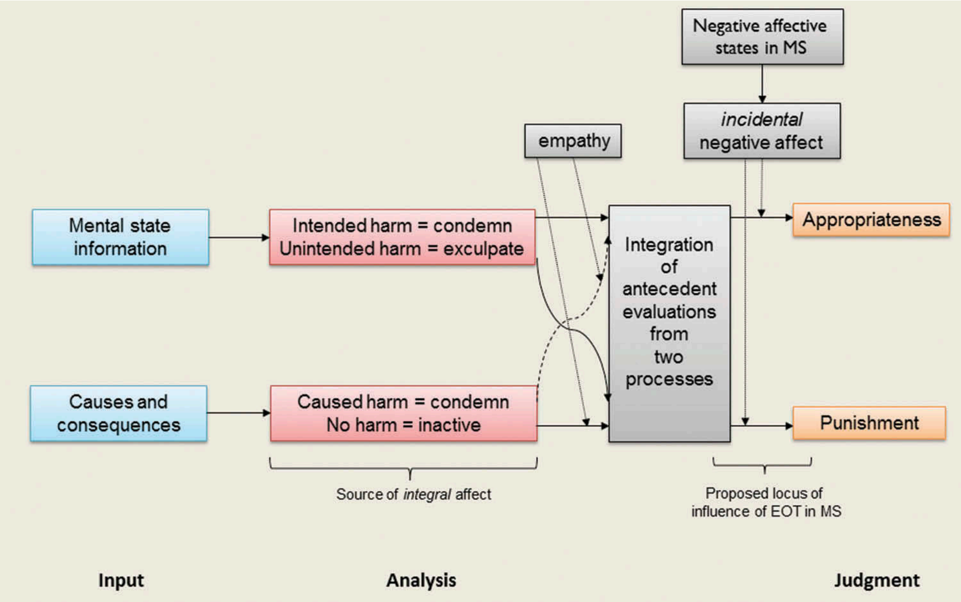

Negative affective states in MS
incidental negative affect
empathy
Mental state information
Intended harm = condemn
Unintended harm = exculpate
Integration of antecedent evaluations from two processes
Appropriateness
Causes and consequences
Caused harm = condemn
No harm = inactive
Punishment
Source of integral affect
Proposed locus of influence of EOT in MS
Input
Analysis
Judgment

Figure 5. Schematic depiction of a two-process model for intent-based moral judgment and the extensions we have proposed based on results from the current study. A causal reasoning process determines the causal role of the agent only in the presence of harmful consequence and condemns the agent, but remains silent in the absence of harmful outcome. A mental state reasoning process probes for culpable mental states and condemns the agent when such a state is found. Both of these processes are sources of negative affect *integral* to the moral violation and provide independent evaluations that are integrated to a different degree for different categories of judgments: punishment judgments rely to a greater degree on information about outcomes than appropriates judgments (represented by solid and dotted lines) and the current study proposes that this influence is mediated by empathy. In other words, outcomes matter more for punitive attitudes only to the degree that people engage in representing victim suffering. Furthermore, based on current findings, we propose that the *incidental* negative affect (e.g., stress, depression, etc.) can amplify the severity of the final integrated judgment in case of poor emotional differentiation skills, stemming from externally focused cognition (like in MS), that insulate judgment from incidental affect and ensure reliance on only the affect integral to the situation. Adapted from the model proposed by Cushman (2008).

[2] The current results motivated us to reanalyze findings from a previous study (Gleichgerrcht et al., 2015), which showed that MS patients were less utilitarian on moral dilemmas than healthy controls due to increased emotional reactivity. This analysis showed that the increased emotional reactivity in MS was no longer significantly different than controls once variance associated with EOT (and not DDF or DIF) was controlled for. Furthermore, higher scores on EOT were predictive of more egocentric moral judgments. These results further bolster our claim that the aberrant moral judgments in MS stem from heightened externally oriented cognitive style that prevents them from properly harnessing affective responses for making moral judgments.

also not significant once group differences in EOT were accounted for. One possible explanation for this effect comes from the observation that higher EOT scores indicate reduced ability to engage in symbolic thought (Herbert, Herbert, & Pollatos, 2011), which has been argued to be important for appreciating others' perspectives and even necessary for acquiring the ability to engage in mentalizing (Tuch, 2011). Thus, the increased tendency to engage in externally oriented thinking coupled with the observed difficulties in MS to properly represent others' mental states may lead MS patients to hold a more egocentric moral frame-of-reference and assume that others share their moral views. The same effect has also been observed for moral judgments on moral dilemmas where MS patients exhibited increased egocentric moral judgments (Gleichgerrcht et al., 2015).

4.4. *Cognitive empathy and the influence of moral luck on punishment versus wrongness*

Although there is evidence to support the claim that outcomes influence punishment judgments to a greater extent than wrongness/appropriateness judgments (Cushman, 2008; Cushman et al., 2009, 2013), the exact psychological mechanism that underpins this process remains unexplored. Neurobiological models of punishment posit empathic aversion stemming from victim distress to be a source of aversive reinforcement that motivates punishment with the ultimate goal of changing the agent's behavior (Martin & Cushman, 2016; Seymour, Singer, & Dolan, 2007). Nonetheless, so far, no further distinction has been made as to precisely which component of empathy (cognitive, affective, or motivational) is involved in conveying the influence of harmful outcomes on punishment judgments.

In the current study, MS patients exhibited reduced motivational empathy (i.e., EC) and increased affective sharing (i.e., PD) as compared to HC, and yet they did not differ from the control population in terms of punishment/wrongness distinction for moral luck contrast cases. In other words, both MS patients and HC increased punishment for accidents (versus neutral cases) more than they increased wrongness judgments. This led us to hypothesize that the cognitive component of empathy (PT), intact in MS patients, may subserve the differential influence of outcomes on punishment versus wrongness judgments (see Figure 5). Accordingly, in the current sample, once the interindividual variation in PT was accounted for, we no longer observed the punishment/wrongness distinction for the key cases. We additionally replicated this effect in a large independent sample and with an expanded set of stimuli in a healthy population from a different cultural setting (see Appendix). This study also showed that removing the variance associated with only PT, and not EC or PD, led to the disappearance of interaction between nature of outcome (harmful or not) and type of moral judgment (punishment or wrongness).

Thus, based on the current findings, we propose a refinement of the existing two-process model for intent-based judgments by showing that cognitive empathy or victim PT is the precise source of affect that motivates a third-party judge to punish agents for causing harmful outcomes. We illustrate the proposed modifications to the existing two-process model in Figure 5.

4.5. *Implications*

The clinical meaningfulness of the current findings requires further exploration, but we argue here that this peculiar pattern of moral judgments can have implications for proper social functioning in MS. Note that not only did MS patients judge others' behavior to be more wrong (or less appropriate) and were more punitive than controls, but they were also more confident that others would evaluate moral situations the same as they do. Although we have provided evidence here only for elevated third-party punishment judgments, we would also expect the same pattern to hold for second-party punishments, more common in day-to-day life, as these judgments recruit similar underlying neural processes (Buckholtz et al., 2008; Yu et al., 2015). This moral profile increases the risk for social dysfunction in MS patients as the individuals interacting with them (their social support system, for instance) may be put off by their punitive attitudes (especially in neutral cases where agents neither intend nor cause harm). In turn, such interpersonal conflict may be exacerbated due to MS patients' apparent inability to appreciate others' moral views during such moral disagreements. This dynamic may lead to greater social exclusion in MS patients, with negative consequences for psychological and social well-being (Montel & Bungener, 2007; Phillips et al., 2009). Thus, intervention strategies designed to improve quality of life in MS patients by increasing their social participation should focus on bringing patients' exaggerated punitiveness to their notice and also encouraging them to see moral situations from others' point-of-view during moral disagreements. This may be achieved by targeting specific socioaffective and sociocognitive processes, such as working on attenuating tendencies to engage in EOT to regulate negative affective states and adopt more adaptive strategies like voluntary attention to emotions (Boden & Thomson, 2015). Additionally, it is known that being aware of the source of the negative affective states can limit its influence on judgments (Västfjäll et al., 2016). Thus, intervention focused on increasing MS patients' awareness of their affective states can make them distrust using this affect as information and reduce the severity of their moral condemnation. This prediction stemming from the current study can be investigated in future studies.

4.6. Limitations

4.6.1. Methodological

The conclusions drawn from this study need to be interpreted with the following limitations in mind. First, as noted in the Methods section, the scenarios were not randomized across conditions, and thus we could not assess the validity of the results from the subjects-level analysis of moral judgments in an item-level analysis to ascertain that main results were not being driven by few artifactual items (cf. McGuire, Langdon, Coltheart, & Mackenzie, 2009). Second, we relied heavily on the IRI questionnaire to assess the different components of empathy, but the IRI has been argued to be a measure with limited ecological validity (Ickes, 2009), and the robustness of the current results needs to be demonstrated with other available instruments that also tap into these different components (e.g., QCAE: Questionnaire of Cognitive and Affective Empathy; Reniers, Corcoran, Drake, Shryane, & Völlm, 2011). Third, the current study did not directly assess ToM deficits in the MS population, but we do note the substantial literature demonstrating these deficits in MS patients (Banati et al., 2010; Charvet et al., 2014; Henry et al., 2009, 2011; Kraemer et al., 2013; Ouellet et al., 2010; Pöttgen et al., 2013; Roca et al., 2014). Fourth, this was an exploratory study that found EOT to play an important role in moral cognition, and thus future studies need to further explore the exact mechanism by which this dimension of alexithymia interacts with incidental affect and how this interaction has downstream consequences for moral judgments. Fifth, the Spanish version of the TAS-20 used in the current study was validated in the general population (Martínez Sánchez, 1996) but not in the MS population. One recent study (Fernández-Jiménez et al., 2013) assessed the factor structure of TAS in the MS population, and future studies should use this version to study cognition in MS. Lastly, it is important to keep in mind that these results are statistical in nature and are not the case reports of profound social dysfunction and moral deficits associated with brain pathology in the literature. This limitation is also important in the light of work that reveals how people usually exhibit discrepancy between judgments they provide on hypothetical scenarios and their behavior in a more ecologically valid setting (Patil, Cogoni, Zangrando, Chittaro, & Silani, 2014). Hence, the translation of the findings will need a clear correlation to real-world functioning parameters.

4.6.2. Statistical

Note that our main conclusions are based on ANCOVAs with covariates that differed between groups (e.g., EOT; in Sections 3.4, 3.5, and 3.7). This has been argued to be a weak procedure as there is no meaningful way to adjust for covariates when they share significant variance with the group factor; in such cases, removing the variance associated with the covariate also alters the grouping variable substantively (Miller & Chapman, 2001). But note that in the current study, these ANCOVAs were carried out exclusively to *explore* the source of the observed effect and not as *hypothesis testing*, a specific case where this practice has been considered to be defensible (Huitema, 2011, p. 212). We did not a priori predict either increased punitiveness or moral egocentricism, nor did we have any prediction for a possible role of EOT in explaining this effect; ANCOVA was carried out to explore the possible role of different independent variables with clinical relevance to the MS population. The current findings should be confirmed in a future hypothesis-driven (rather than exploratory) study where ANCOVA can be circumvented by matching the MS group for EOT scores with HC. In such a study, the current proposal would predict that the MS patients will not exhibit harsher moral condemnation profile.

5. Conclusions

In the current study, we investigated intent-based moral judgments in MS patients to assess how the observed deficits in social cognition in this population affect their moral evaluations. We found that although MS patients judged specific types of third-party moral violations appropriately by integrating information about beliefs and outcomes, they nonetheless judged behaviors to be *generally* more wrong, they were more punitive, and they were more confident about their judgments as compared to HC. We argued building on past work that this increased moral condemnation in MS is likely to be due to unrestrained influence of prevalent, *incidental* negative affective states whose influence on moral judgments is not differentiated from *integral* affect stemming from stimuli due to externally oriented cognitive style. Moreover, more egocentric moral attitudes in MS patients may also arise from the confluent effect of self-centered perspective due to external preoccupations and reduced PT due to ToM deficits. The distinction between wrongness and punishment judgments for moral luck cases was found to be preserved in MS patients due to an intact cognitive capacity for victim PT. Finally, we point out the translational value of the current research for therapeutic practices designed to improve quality of life in MS patients.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Indrajeet Patil http://orcid.org/0000-0003-1995-6531
Ezequiel Gleichgerrcht http://orcid.org/0000-0002-4212-4146

References

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716. doi:10.1146/annurev.psych.60.110707.163514

Albiero, P., Ingoglia, S., & Lo Coco, A. (2006). Contributo all'adattamento italiano dell'Interpersonal Reactivity Index [A contribution to the Italian validation of the Interpersonal Reactivity Index]. *Testing Psicometria Metodologia*, *13*(2), 107–125.

Alter, A. L., Kernochan, J., & Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law and Human Behavior*, *31*(4), 319–335. doi:10.1007/s10979-006-9060-x

Baez, S., Couto, B., Torralva, T., Sposato, L. A., Huepe, D., Montañes, P., … Ibanez, A. (2014). Comparing moral judgments of patients with frontotemporal dementia and frontal stroke. *JAMA Neurology*, *71*(9), 1172–1176. doi:10.1001/jamaneurol.2014.347

Baez, S., Kanske, P., Matallana, D., Montanes, P., Reyes, P., Vigliecca, N. S., & Ibanez, A. (2016). Integration of intention and outcome for moral judgment in frontotemporal dementia: Brain structural signatures. *Neurodegenerative Diseases*. doi:10.1159/000441918

Baez, S., Manes, F., Huepe, D., Torralva, T., Fiorentino, N., Richter, F., … Ibanez, A. (2014). Primary empathy deficits in frontotemporal dementia. *Frontiers in Aging Neuroscience*, *6*, 262. doi:10.3389/fnagi.2014.00262

Baez, S., Morales, J. P., Slachevsky, A., Torralva, T., Matus, C., Manes, F., & Ibanez, A. (2016). Orbitofrontal and limbic signatures of empathic concern and intentional harm in the behavioral variant frontotemporal dementia. *Cortex*, *75*, 20–32. doi:10.1016/j.cortex.2015.11.007

Baez, S., Rattazzi, A., Gonzalez-Gadea, M. L., Torralva, T., Vigliecca, N. S., Decety, J., … Ibanez, A. (2012). Integrating intention and context: Assessing social cognition in adults with Asperger syndrome. *Frontiers in Human Neuroscience*, *6*, 302. doi:10.3389/fnhum.2012.00302

Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994). The Twenty-Item Toronto Alexithymia Scale: II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research*, *38*, 33–40. doi:10.1016/0022-3999(94)90006-X

Baird, J., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *103*, 37–49. doi:10.1002/cd.96

Baldner, C., & McGinley, J. J. (2014). Correlational and exploratory factor analyses (EFA) of commonly used empathy questionnaires: New insights. *Motivation and Emotion*, *38*(5), 727–744. doi:10.1007/s11031-014-9417-2

Banati, M., Sandor, J., Mike, A., Illes, E., Bors, L., Feldmann, A., … Illes, Z. (2010). Social cognition and Theory of Mind in patients with relapsing-remitting multiple sclerosis. *European Journal of Neurology*, *17*(3), 426–433. doi:10.1111/j.1468-1331.2009.02836.x

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1522070113

Beatty, W. W., Orbelo, D. M., Sorocco, K. H., & Ross, E. D. (2003). Comprehension of affective prosody in multiple sclerosis. *Multiple Sclerosis*, *9*(2), 148–153. doi:10.1191/1352458503ms897oa

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*(7), e285. doi:10.1038/tp.2013.61

Boden, M. T., & Thompson, R. J. (2015). Facets of emotional awareness and associations with emotion regulation and depression. *Emotion*, *15*(3), 399–410.

Bodini, B., Mandarelli, G., Tomassini, V., Tarsitani, L., Pestalozza, I., Gasperini, C., … Pozzilli, C. (2008). Alexithymia in multiple sclerosis: Relationship with fatigue and depression. *Acta Neurologica Scandinavica*, *118*(1), 18–23. doi:10.1111/j.1600-0404.2007.00969.x

Bonavita, S., Tedeschi, G., & Gallo, A. (2013). Morphostructural MRI abnormalities related to neuropsychiatric disorders associated to multiple sclerosis. *Multiple Sclerosis International*, *2013*, 102454. doi:10.1155/2013/102454

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60*(5), 930–940. doi:10.1016/j.neuron.2008.10.016

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., & Marois, R. (2015). From blame to punishment: Disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron*, *87*(6), 1369–1380. doi:10.1016/j.neuron.2015.08.023

Buon, M., Dupoux, E., Jacob, P., Chaste, P., Leboyer, M., & Zalla, T. (2013). The role of causal and intentional judgments in moral reasoning in individuals with high functioning autism. *Journal of Autism and Developmental Disorders*, *43*(2), 458–470. doi:10.1007/s10803-012-1588-7

Buon, M., Jacob, P., Loissel, E., & Dupoux, E. (2013). A non-mentalistic cause-based heuristic in human social evaluations. *Cognition*, *126*(2), 149–155. doi:10.1016/j.cognition.2012.09.006

Calabrese, M., Rinaldi, F., Mattisi, I., Grossi, P., Favaretto, A., Atzori, M., … Gallo, P. (2010). Widespread cortical thinning characterizes patients with MS with mild cognitive impairment. *Neurology*, *74*(4), 321–328. doi:10.1212/WNL.0b013e3181cbcd03

Cameron, C. D., Payne, B. K., & Doris, J. M. (2013). Morality in high definition: Emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of Experimental Social Psychology*, *49*(4), 719–725. doi:10.1016/j.jesp.2013.02.014

Cecchetto, C., Aiello, M., D'Amico, D., Cutuli, D., Cargnelutti, D., Eleopra, R., & Rumiati, R. I. (2014). Facial and bodily emotion recognition in multiple sclerosis: The role of alexithymia and other characteristics of the disease. *Journal of the International Neuropsychological Society*, *20*(10), 1004–1014. doi:10.1017/S1355617714000939

Chahraoui, K., Duchene, C., Rollot, F., Bonin, B., & Moreau, T. (2014). Longitudinal study of alexithymia and multiple sclerosis. *Brain and Behavior*, *4*(1), 75–82. doi:10.1002/brb3.194

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, *11*(3), 476–484. doi:10.1093/scan/nsv131

Charvet, L. E., Cleary, R. E., Vazquez, K., Belman, A. L., & Krupp, L. B. (2014). Social cognition in pediatric-onset multiple sclerosis (MS). *Multiple Sclerosis Journal*, *20*(11), 1478–1484. doi:10.1177/1352458514526942

Cheng, J. S., Ottati, V. C., & Price, E. D. (2013). The arousal model of moral condemnation. *Journal of Experimental Social Psychology*, *49*(6), 1012–1018. doi:10.1016/j.jesp.2013.06.006

Ciaramelli, E., Braghittoni, D., & di Pellegrino, G. (2012). It is the outcome that counts! Damage to the ventromedial prefrontal cortex disrupts the integration of outcome and belief information for moral judgment. *Journal of the International Neuropsychological Society*, *18*, 962–971. doi:10.1017/S1355617712000690

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. doi:10.1016/j.cognition.2008.03.006

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, *6*, 97–103. doi:10.1016/j.copsyc.2015.06.003

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PloS one*, *4*(8), e6699. doi:10.1371/journal.pone.0006699

Cushman, F. A., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*(1), 6–21. doi:10.1016/j.cognition.2012.11.008

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. doi:10.1037/0022-3514.44.1.113

Davydov, D. M., Luminet, O., & Zech, E. (2013). An externally oriented style of thinking as a moderator of responses to affective films in women. *International Journal of Psychophysiology*, *87*(2), 152–164. doi:10.1016/j.ijpsycho.2012.12.003

Davydov, D. M., Stewart, R., Ritchie, K., & Chaudieu, I. (2010). Resilience and mental health. *Clinical Psychology Review*, *30*(5), 479–495. doi:10.1016/j.cpr.2010.03.003

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. doi:10.1016/j.tics.2006.08.008

Decety, J., & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology*, *108*(11), 3068–3072. doi:10.1152/jn.00473.2012

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, *9*(5), 525–537. doi:10.1177/1745691614545130

Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral Cortex*, *22*, 209–220. doi:10.1093/cercor/bhr111

Decety, J., & Yoder, K. J. (2015). Empathy and motivation for justice: Cognitive empathy and concern, but not emotional empathy, predict sensitivity to injustice for others. *Social Neuroscience*, *11*(1), 1–14. doi:10.1080/17470919.2015.1029593

Demers, L. A., & Koven, N. S. (2015). The relation of alexithymic traits to affective theory of mind. *The American Journal of Psychology*, *128*(1), 31–42. doi:10.5406/amerjpsyc.128.1.0031

Di Schiena, R., Luminet, O., & Philippot, P. (2011). Adaptive and maladaptive rumination in alexithymia and their relation with depressive symptoms. *Personality and Individual Differences*, *50*(1), 10–14. doi:10.1016/j.paid.2010.07.037

Evren, C., Cınar, O., & Evren, B. (2012). Relationship of alexithymia and dissociation with severity of borderline personality features in male substance-dependent inpatients. *Comprehensive Psychiatry*, *53*(6), 854–859. doi:10.1016/j.comppsych.2011.11.009

Feinstein, A., Magalhaes, S., Richard, J. F., Audet, B., & Moore, C. (2014). The link between multiple sclerosis and depression. *Nature Reviews Neurology*, *10*(9), 507–517. doi:10.1038/nrneurol.2014.139

Fernández-Jiménez, E., Pérez-San-Gregorio, M. A., Taylor, G. J., Bagby, R. M., Ayearst, L. E., & Izquierdo, G. (2013). Psychometric properties of a revised Spanish 20-item Toronto Alexithymia Scale adaptation in multiple sclerosis patients. *International Journal of Clinical and Health Psychology*, *13*, 226–234. doi:10.1016/S1697-2600(13)70027-9

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. doi:10.1016/0022-3956(75)90026-6

Fox, R. J. (2008). Picturing multiple sclerosis: Conventional and diffusion tensor imaging. *Seminars in Neurology*, *28*(4), 453–466. doi:10.1055/s-0028-1083689

Gan, T., Lu, X., Li, W., Gui, D., Tang, H., Mai, X., & Luo, Y. J. (2015). Temporal dynamics of the integration of intention and outcome in harmful and helpful moral judgment. *Frontiers in Psychology*, *6*, 2022. doi:10.3389/fpsyg.2015.02022

Gay, M. C., Vrignaud, P., Garitte, C., & Meunier, C. (2010). Predictors of depression in multiple sclerosis patients. *Acta Neurologica Scandinavica*, *121*(3), 161–170. doi:10.1111/j.1600-0404.2009.01232.x

Gleichgerrcht, E., Tomashitis, B., & Sinay, V. (2015). The relationship between alexithymia, empathy and moral judgment in patients with multiple sclerosis. *European Journal of Neurology*, *22*(9), 1295–1303. doi:10.1111/ene.12745

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PloS one*, *8*(4), e60418. doi:10.1371/journal.pone.0060418

Güleç, M. Y., Altintaş, M., İnanç, L., Bezgin, Ç. H., Koca, E. K., & Güleç, H. (2013). Effects of childhood trauma on somatization in major depressive disorder: The role of alexithymia. *Journal of Affective Disorders*, *146*(1), 137–141. doi:10.1016/j.jad.2012.06.033

Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second-and third-party punishment behavior. *Cognition*, *133*(1), 97–103. doi:10.1016/j.cognition.2014.06.001

Hendryx, M. S., Haviland, M. G., & Shaw, D. G. (1991). Dimensions of alexithymia and their relationships to anxiety and depression. *Journal of Personality Assessment*, *56*(2), 227–237. doi:10.1207/s15327752jpa5602_4

Henry, A., Tourbah, A., Chaunu, M. P., Rumbach, L., Montreuil, M., & Bakchine, S. (2011). Social cognition impairments in relapsing-remitting multiple sclerosis. *Journal of the International Neuropsychological Society*, *17*(06), 1122–1131. doi:10.1017/S1355617711001147

Henry, J. D., Phillips, L. H., Beatty, W. W., McDonald, S., Longley, W. A., Joscelyne, A., & Rendell, P. G. (2009). Evidence for deficits in facial affect recognition and theory of mind in multiple sclerosis. *Journal of the International Neuropsychological Society*, *15*(2), 277–285. doi:10.1017/S1355617709090195

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the relationship between interoceptive awareness and alexithymia: Is interoceptive awareness related to emotional awareness? *Journal of Personality*, *79*(5), 1149–1175. doi:10.1111/j.1467-6494.2011.00717.x

Hesse, E., Mikulan, E., Decety, J., Sigman, M., Garcia, M. D., Silva, W., … Ibanez, A. (2016). Early detection of intentional harm in the human amygdala. *Brain*, *139*(1), 54–61. doi:10.1093/brain/awv336

Huitema, B. E. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies* (2nd ed.). John Wiley & Sons, Ltd. doi:10.1002/9781118067475

Ickes, W. (2009). Empathic accuracy: Its links to clinical, cognitive, developmental, social, and physiological psychology. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 57–70). Cambridge, MA: The MIT Press.

Jehna, M., Langkammer, C., Wallner-Blazek, M., Neuper, C., Loitfelder, M., Ropele, S., … Enzinger, C. (2011). Cognitively preserved MS patients demonstrate functional differences in processing neutral and emotional faces. *Brain Imaging and Behavior*, *5*(4), 241–251. doi:10.1007/s11682-011-9128-1

Koster-Hale, J., & Saxe, R. (2013). Functional neuroimaging of theory of mind. In S. Baron-Cohen, M. Lombardo, & H. Tager-Flusberg (Eds.), *Understanding other minds* (3rd ed., pp. 132–163). Oxford: Oxford University Press.

Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences*, *110*(14), 5648–5653. doi:10.1073/pnas.1207992110

Koven, N. S. (2011). Specificity of meta-emotion effects on moral decision-making. *Emotion*, *11*(5), 1255–1261. doi:10.1037/a0025616

Kraemer, M., Herold, M., Uekermann, J., Kis, B., Wiltfang, J., Daum, I., … Abdel-Hamid, M. (2013). Theory of mind and empathy in patients at an early stage of relapsing remitting multiple sclerosis. *Clinical Neurology and Neurosurgery*, *115*(7), 1016–1022. doi:10.1016/j.clineuro.2012.10.027

Krause, M., Wendt, J., Dressel, A., Berneiser, J., Kessler, C., Hamm, A. O., & Lotze, M. (2009). Prefrontal function associated with impaired emotion recognition in patients with multiple sclerosis. *Behavioural Brain Research*, *205*(1), 280–285. doi:10.1016/j.bbr.2009.08.009

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444. doi:10.1212/WNL.33.11.1444

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. doi:10.3389/fpsyg.2013.00863

Lakens, D. (2015a, June 8). Why you should use omega-squared instead of eta-squared. *Blog post*. Retrieved from http://daniellakens.blogspot.it/2015/06/why-you-should-use-omega-squared.html

Lakens, D. (2015b, February 27). Which statistics should you report? *Blog post*. Retrieved from http://daniellakens.blogspot.it/2015/02/which-statistics-should-you-report.html

Lakens, D. (2015c, January 26). Always use Welch's t-test instead of Student's t-test. *Blog post*. Retrieved from http://daniellakens.blogspot.it/2015/01/always-use-welchs-t-test-instead-of.html

Lindquist, K., & Barrett, L. F. (2008). Emotional complexity. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *The handbook of emotion* (3rd ed., pp. 513–530). New York, NY: Guilford.

Luminet, O., Rimé, B., Bagby, R. M., & Taylor, G. (2004). A multimodal investigation of emotional responding in alexithymia. *Cognition & Emotion*, *18*(6), 741–766. doi:10.1080/02699930341000275

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186. doi:10.1080/1047840X.2014.877340

Marchesi, C., Brusamonti, E., & Maggini, C. (2000). Are alexithymia, depression, and anxiety distinct constructs in affective disorders? *Journal of Psychosomatic Research*, *49*(1), 43–49. doi:10.1016/S0022-3999(00)00084-2

Martin, J. W., & Cushman, F. A. (2016). The adaptive logic of moral luck. In J. Sytsma & W. Buckwalter (Eds.), *The Blackwell companion to experimental philosophy* (pp. 190–202). Wiley Blackwell.

Martínez Sánchez, F. (1996). Adaptación española de la escala de Alexitimia de Toronto (TAS-20) [Spanish adaptation of the Toronto Alexithymia Scale (TAS-20)]. *Clínica y Salud*, *7* (1), 19–32.

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, *45*(3), 577–580. doi:10.1016/j.jesp.2009.01.002

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*, 40–48. doi:10.1037/0021-843X.110.1.40

Monson, C. M., Price, J. L., Rodriguez, B. F., Ripley, M. P., & Warner, R. A. (2004). Emotional deficits in military-related PTSD: An investigation of content and process disturbances. *Journal of Traumatic Stress*, *17*(3), 275–279. doi:10.1023/B:JOTS.0000029271.58494.05

Montel, S. R., & Bungener, C. (2007). Coping and quality of life in one hundred and thirty five subjects with multiple sclerosis. *Multiple Sclerosis*, *13*(3), 393–401. doi:10.1177/1352458506071170

Moran, J., Young, L., Saxe, R., Lee, S., O'Young, D., Mavros, P., & Gabrieli, J. (2011). Impaired theory of mind for moral judgment in high functioning autism. *Proceedings of the National Academy of Sciences*, *108*(7), 2688–2692. doi:10.1073/pnas.1011734108

Moriguchi, Y., & Komaki, G. (2013). Neuroimaging studies of alexithymia: Physical, affective, and social perspectives. *BioPsychoSocial Medicine*, *7*, 8. doi:10.1186/1751-0759-7-8

Nemiah, J. C., Freyberger, H., & Sifneos, P. E. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (ed.), *Modern trends in psychosomatic medicine* (pp. 430–439). London: Butterworths.

Ngo, L., Kelly, M., Coutlee, C. G., Carter, R. M., Sinnott-Armstrong, W., & Huettel, S. A. (2015). Two distinct moral mechanisms for ascribing and denying intentionality. *Scientific Reports*, *5*, 17390. doi:10.1038/srep17390

Nigro, S., Passamonti, L., Riccelli, R., Toschi, N., Rocca, F., Valentino, P., … Quattrone, A. (2015). Structural 'connectomic'alterations in the limbic system of multiple sclerosis patients with major depression. *Multiple Sclerosis Journal*, *21*(8), 1003–1012. doi:10.1177/1352458514558474

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini-review. *Frontiers in Psychology*, *3*, 322. doi:10.3389/fpsyg.2012.00322

Ouellet, J., Scherzer, P. B., Rouleau, I., Metras, P., Bertrand-Gauvin, C., Djerroud, N., … Duquette, P. (2010). Assessment of social cognition in patients with multiple sclerosis. *Journal of the International Neuropsychological Society*, *16*(2), 287–296. doi:10.1017/S1355617709991329

Passamonti, L., Cerasa, A., Liguori, M., Gioia, M. C., Valentino, P., Nisticò, R., … Fera, F. (2009). Neurobiological mechanisms underlying emotional processing in relapsing-remitting multiple sclerosis. *Brain*, *132*(12), 3380–3391. doi:10.1093/brain/awp095

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. doi:10.1080/17470919.2013.870091

Patil, I., Melsbach, J., Hennig-Fast, K., & Silani, G. (2016). Divergent roles of autistic and alexithymic traits in utilitarian moral judgments in adults with autism. *Scientific Reports*, *6*, 23637. doi:10.1038/srep23637

Patil, I., & Silani, G. (2014b). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*, e501. doi:10.3389/fpsyg.2014.00501

Patil, I., & Silani, G. (2014a). Alexithymia increases moral acceptability of accidental harms. *Journal of Cognitive Psychology*, *26*(5), 597–614. doi:10.1080/20445911.2014.929137

Pepping, M., Brunings, J., & Goldberg, M. (2013). Cognition, cognitive dysfunction, and cognitive rehabilitation in multiple sclerosis. *Physical Medicine and Rehabilitation Clinics of North America*, *24*(4), 663–672. doi:10.1016/j.pmr.2013.06.009

Pérez-Albéniz, A., De Paúl, J., Etxeberría, J., Montes, M. P., & Torres, E. (2003). Adaptación de interpersonal reactivity index (IRI) al español [Adaptation of Interpersonal Reactivity Index (IRI) in Spanish]. *Psicothema*, *15*(2), 267–272.

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2012). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, *3*, 606. doi:10.3389/fpsyg.2012.00606

Phillips, L. H., Saldias, A., McCarrey, A., Henry, J. D., Scott, C., Summers, F., & Whyte, M. (2009). Attentional lapses, emotional regulation and quality of life in multiple sclerosis. *British Journal of Clinical Psychology*, *48*(1), 101–106. doi:10.1348/014466508X379566

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., … Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology*, *69*(2), 292–302. doi:10.1002/ana.22366

Pöttgen, J., Dziobek, I., Reh, S., Heesen, C., & Gold, S. M. (2013). Impaired social cognition in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(5), 523–528. doi:10.1136/jnnp-2012-304157

Prehn, K., Wartenburger, I., Mériau, K., Scheibe, C., Goodenough, O. R., Villringer, A., … Heekeren, H. R. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience*, *3*(1), 33–46. doi:10.1093/scan/nsm037

Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84–95. doi:10.1080/00223891.2010.528484

Roca, M., Manes, F., Gleichgerrcht, E., Ibáñez, A., de Toledo, M. E. G., Marenco, V., … Sinay, V. (2014). Cognitive but not affective theory of mind deficits in mild relapsing-remitting multiple sclerosis. *Cognitive and Behavioral Neurology*, *27*(1), 25–30. doi:10.1097/WNN.0000000000000017

Rocca, M. A., Amato, M. P., De Stefano, N., Enzinger, C., Geurts, J. J., Penner, I.-K., … Filippi, M. (2015). Clinical and imaging assessment of cognitive dysfunction in multiple sclerosis. *The Lancet Neurology*, *14*(3), 302–317.

Roxburgh, R. H. S. R., Seaman, S. R., Masterman, T., Hensiek, A. E., Sawcer, S. J., Vukusic, S., … Compston, D. A. S. (2005). Multiple Sclerosis Severity Score: Using disability and disease duration to rate disease severity. *Neurology*, *64*(7), 1144–1151. doi:10.1212/01.WNL.0000156155.19270.F8

Sá, M. J. (2008). Psychological aspects of multiple sclerosis. *Clinical Neurology and Neurosurgery*, *110*(9), 868–877. doi:10.1016/j.clineuro.2007.10.001

Saarijärvi, S., Salminen, J. K., & Toikka, T. B. (2001). Alexithymia and depression: A 1-year follow-up study in outpatients with major depression. *Journal of Psychosomatic Research*, *51*(6), 729–733. doi:10.1016/S0022-3999(01)00257-4

Sacco, R., Bonavita, S., Esposito, F., Tedeschi, G., & Gallo, A. (2013). The contribution of resting state networks to the study of cortical reorganization in MS. *Multiple Sclerosis International*, *2013*, 857807. doi:10.1155/2013/857807

Salvano-Pardieu, V., Blanc, R., Combalbert, N., Pierratte, A., Manktelow, K., Maintier, C., … Fontaine, R. (2015). Judgment of blame in teenagers with Asperger's syndrome. *Thinking & Reasoning*, 1–23. doi:10.1080/13546783.2015.1127288

Sellaro, R., Güroğlu, B., Nitsche, M. A., van den Wildenberg, W. P. M., Massaro, V., Durieux, J., … Colzato, L. S. (2015). Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction. *Neuropsychologia*, *77*, 400–408. doi:10.1016/j.neuropsychologia.2015.09.016

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience*, *8*(4), 300–311. doi:10.1038/nrn2119

Sloman, S. A., Fernbach, P. M., & Ewing, S. (2009). Causal models: The representational infrastructure for moral judgement. In D. Bartels, C. W. Bauman, L. J. Skitka, & D. Medin (Eds.), *Moral judgment and decision making: The psychology of learning and motivation* (Vol. 50). San Diego, CA: Elsevier.

Treadway, M. T., Buckholtz, J. W., Martin, J. W., Jan, K., Asplund, C. L., Ginther, M. R., … Marois, R. (2014). Corticolimbic gating of emotion-driven punishment. *Nature Neuroscience*, *17*(9), 1270–1275. doi:10.1038/nn.3781

Trémolière, B., & Djeriouat, H. (2016). The sadistic trait predicts minimization of intention and causal responsibility in moral judgment. *Cognition*, *146*, 158–171. doi:10.1016/j.cognition.2015.09.014

Tuch, R. H. (2011). Thinking outside the box: A metacognitive / theory of mind perspective on concrete thinking. *Journal of the American Psychoanalytic Association*, *59*(4), 765–789. doi:10.1177/0003065111417625

Västfjäll, D., Slovic, P., Burns, W. J., Erlandsson, A., Koppel, L., Asutay, E., & Tinghög, G. (2016). The arithmetic of emotion: Integration of incidental and integral affect in judgments and decisions. *Frontiers in Psychology*, *7*, 325. doi:10.3389/fpsyg.2016.00325

Walter, N. T., Montag, C., Markett, S., Felten, A., Voigt, G., & Reuter, M. (2012). Ignorance is no excuse: Moral judgments are influenced by a genetic variation on the oxytocin receptor gene. *Brain and Cognition*, *78*(3), 268–273. doi:10.1016/j.bandc.2012.01.003

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLoS Biology*, *13*(4), e1002128. doi:10.1371/journal.pbio.1002128

Ye, H., Chen, S., Huang, D., Zheng, H., Jia, Y., & Luo, J. (2015). Modulation of neural activity in the temporoparietal junction with transcranial direct current stimulation changes the role of beliefs in moral judgment. *Frontiers in Human Neuroscience*, *9*, 659. doi:10.3389/fnhum.2015.00659

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*, 845–851. doi:10.1016/j.neuron.2010.03.003

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*, 6753–6758. doi:10.1073/pnas.0914826107

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240. doi:10.1073/pnas.0701408104

Young, L., Koenigs, M., Kruepke, M., & Newman, J. (2012). Psychopathy increases perceived moral permissibility of accidents. *Journal of Abnormal Psychology*, *121*(3), 659–667. doi:10.1037/a0027489

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*(4), 1912–1920.

Young, L., & Saxe, R. (2009). Innocent Intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*, 2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*, *7*(8), 585–604.

Yu, H., Li, J., & Zhou, X. (2015). Neural substrates of intention–consequence integration and its impact on reactive punishment in interpersonal transgression. *The Journal of Neuroscience*, *35*(12), 4917–4925. doi:10.1523/JNEUROSCI.3536-14.2015

Zackheim, L. (2007). Alexithymia: The expanding realm of research. *Journal of Psychosomatic Research*, *63*(4), 345–347. doi:10.1016/j.jpsychores.2007.08.011

Appendix

Replication study

Participants

A total of 113 healthy native Italian-speaking community members (70 females) came to the lab to participate in this study and were financially compensated for their time. All participants provided written informed consent. Average age was 24.42 (SD = 5.55, median = 23) years, with a range of 18–57.

Experimental stimuli

Experimental stimuli consisted of four different sets or versions of 36 unique vignettes for a total of 144 stories and participants were assigned different versions in Latin-square arrangement. Each participant saw one variation of each scenario, for a total of 36 stories. All scenarios were adapted in Italian from Young, Camprodon, et al. (2010). The four versions were the result of a 2-by-2 design where the factors *belief* (neutral, negative) and *outcome* (neutral, negative) were independently varied such that agents in the scenario produced either a neutral outcome or a harmful outcome while acting with the belief that they were causing either a neutral outcome or a harmful outcome (see Figure A1).

Each scenario lasted for 32 s and consisted of four cumulative segments (each lasting for 8 s): (i)

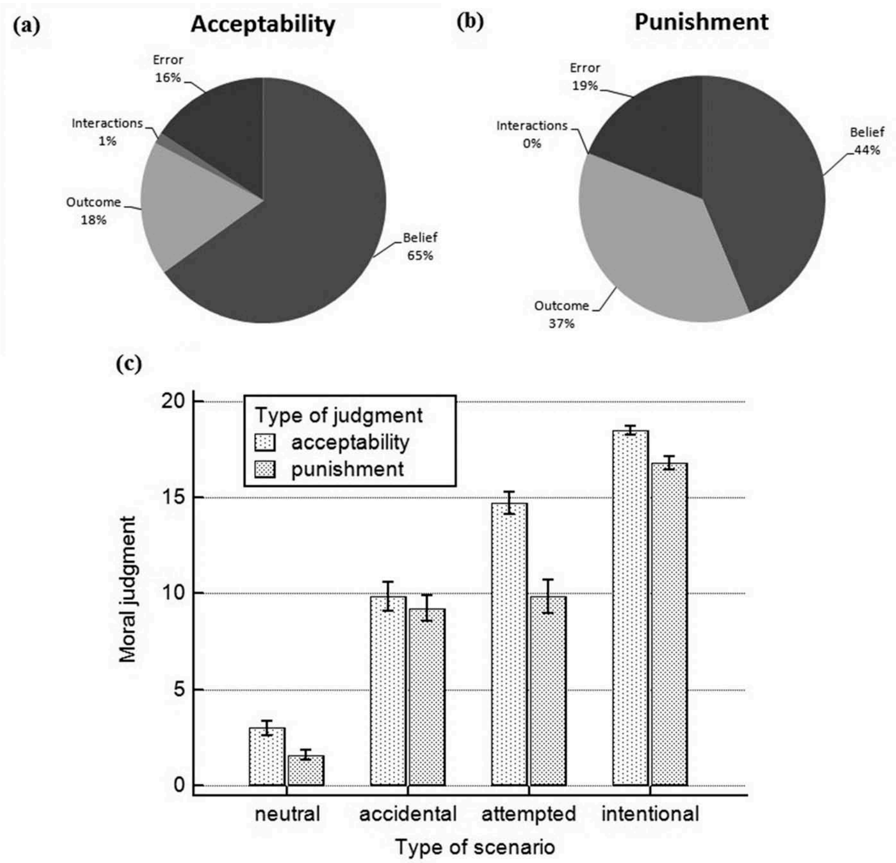


Figure A1. Results from the replication study. (*a*) Proportion of within-condition variability explained by each factor for the acceptability judgment. (*b*) Proportion of within-condition variability explained by each factor for the punishment judgment. (*c*) Direct comparison of mean moral judgment for each condition grouped by type of judgment.

background: this stem was common to all variations and provided settings in which the story took place; (ii) *foreshadow*: this segment foreshadowed whether the outcome will be neutral or harmful; (iii) *belief*: this segment provided information about whether the agent was acting with a neutral or harmful belief; (iv) *consequence*: this final segment revealed the outcome of the agent's action. All the story text was then removed and replaced with the question and response scale.

After reading each scenario, participants provided two types of moral judgments, which were presented in a randomized order:

(i) *acceptability*: "How morally acceptable was [the agent]'s behavior?" (from "Not at all acceptable" to "Completely acceptable");
(ii) *punishment*: "How much punishment does [the agent] deserve for his/her behavior?" (from "No punishment" to "The most severe punishment").

After each story was presented, participants responded using computerized visual analog scales (VAS), implemented as horizontal onscreen bar and responses were later converted to standardized scores with [min, max] of [0, 20]. The acceptability scores were reverse-scored so that higher score for the two questions indicated less acceptable behavior and more punishment. Participants had 6 s to respond to each question.

Participants then completed the Italian-validated version of the IRI (Albiero et al., 2006). Note that all subscales of the Italian version had the same items as that in the English version.

Data analysis

All effect sizes have been reported based on recommendations in Lakens (2013).

Results

For the acceptability judgments, the belief and outcome factors, respectively, accounted for 65% and 18% of the variability (Figure A1(a)), while for the punishment judgments these factors accounted for 44% and 37% of the variance (Figure A1(b)). Thus, as predicted by the two-process model, punishment judgments were dependent to an equal degree on both belief and outcome information, while the acceptability judgments depended to a large extent on the mental state information.

ANOVA analysis also revealed the same pattern: There was a main effect of belief and outcome for both acceptability (belief: $F(1,112) = 835.08$, $p < 0.001$, $p\eta^2 = 0.882$, 90% CI [0.849, 0.902]; outcome: $F(1,112) = 448.63$, $p < 0.001$, $p\eta^2 = 0.800$, 90% CI [0.746, 0.835]) and punishment (belief: $F(1,112) = 516.77$, $p < 0.001$, $p\eta^2 = 0.822$, 90% CI [0.773, 0.853]; outcome: $F(1,112) = 714.43$, $p < 0.001$, $p\eta^2 = 0.864$, 90% CI [0.827, 0.888]) judgments such that behavior of agents who produced harmful outcome or who harbored intention to hard was condemned and punished more severely (see Figure A1(c)). Additionally, there was also a belief-by-outcome interaction for acceptability ($F(1,112) = 62.80$, $p < 0.001$, $p\eta^2 = 0.359$, 90% CI [0.243, 0.456]) but not for punishment ($F(1,112) = 2.71$, $p = 0.102$) judgments.

More important to the current replication study, we observed an outcome-by-type of judgment interaction in a 2 (belief) × 2 (outcome) × 2 (type of judgment) repeated measures ANOVA ($F(1,112) = 81.893$, $p < 0.001$, $p\eta^2 = 0.422$). This interaction signifies that participants relied on outcome information to a different degree while evaluating acceptability of and punishment for the behavior under consideration. In particular, punishment for behavior of an agent with harmful intent was *reduced* significantly more than acceptability when she failed as compared to when she succeeded in producing harmful outcome ($F(1,112) = 88.29$, $p < 0.001$, $p\eta^2 = 0.441$, 90% CI [0.327, 0.530]). Similarly, severity of endorsed punishment for an agent who accidently produced harm while acting under false belief was *increased* significantly more than acceptability as compared to a neutral case ($F(1,112) = 10.28$, $p = 0.002$, $p\eta^2 = 0.084$, 90% CI [0.020, 0.173]).

To address the primary replication study hypothesis, we reran the 2 (belief) × 2 (outcome) × 2 (type of judgment) repeated-measures ANOVA but with PT as a covariate. This model did not reveal any significant interaction between outcome and judgment type ($F(1,111) = 2.393$, $p = 0.125$, $p\eta^2 = 0.021$). But this interaction remained significant if EC ($F(1,111) = 6.016$, $p = 0.016$, $p\eta^2 = 0.051$) or PD ($F(1,111) = 7.355$, $p = 0.008$, $p\eta^2 = 0.064$) was added as a covariate to the model.

Thus, we replicated our finding from the main study that once the interindividual differences in cognitive empathy are accounted for, outcomes no longer have more influence on punishment judgments than the wrongness judgments. In other words, perspective-taking, and not empathic concern or PD (which are motivational and affective aspects of empathy, respectively), seems to be at the core of why moral luck has a greater bearing on punishment judgments than wrongness judgments. This also explains why MS patients showed the usual wrongness/punishment distinction in terms of reliance on outcome information; although MS patients showed reduced EC and increased PD, they did not differ from the control population in their capacity to engage in cognitive empathizing.