UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI PSICOLOGIA GENERALE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE

XXVIII ciclo

THE ROLE OF EMOTION AND INTENTIONALITY IN MORAL DILEMMA RESOLUTION: SUBJECTIVE, BEHAVIORAL, AND ELECTROPHYSIOLOGICAL DATA

Direttore della Scuola: Prof.ssa Francesca Peressotti

Supervisore: Prof.ssa Michela Sarlo

Dottorando: Carolina Pletti

TABLE OF CONTENTS

1. General Introduction 7

1.1. Emotion and decision-making 7

1.1.1. The somatic marker hypothesis 8

1.1.2. Decision-related emotions 11

1.2. Emotion and morality 13

1.2.1. Intuitive processes in moral judgment 13

1.2.2. Moral dilemmas and the dual process model of moral judgment 15

1.2.3. What makes a moral dilemma a Footbridge-type dilemma? Dilemmas characteristics and categorization 27

1.2.4. The role of harm aversion and intentionality in causing emotional reactions in the moral domain 31

1.3. Aims of the research project and outline of the studies 37

2. Study I: The influence of legal consequences on neural activity, emotional experience and decision choice in moral dilemmas 45

2.1. Introduction 45

2.2. Methods 48

2.2.1. Participants 48

2.2.2. Stimuli 49

2.2.3. Procedure 49

2.2.4. Data Collection and Analysis 51

2.3. Results 52

2.3.1. Behavioral data 52
2.3.2. Affective ratings 52
2.3.3. Stimulus-locked ERPs 53
2.3.4. Response-locked MRPs 55
2.4. Discussion 57
2.5. Conclusion 63
3. Study II: The influence of trait psychopathy on judgment and choices in moral dilemmas 65
3.1. Introduction 65
3.2. Methods 69
3.2.1. Participants 69
3.2.2. Stimuli 69
3.2.3. Procedure 70
3.2.4. Statistical analysis 71
3.3. Results 71
3.3.1. Choice of action 71
3.3.2. Moral judgment 72
3.3.3. Emotional reactivity 73
3.3.4. Correlations 74
3.4. Discussion 74
3.5. Conclusion 77
4. Study III: the role of Anticipated emotions in the resolution of moral dilemmas 79

4.1. Introduction 79
4.2. Methods 84
4.2.1. Participants 84
4.2.2. Stimuli 85
4.2.3. Procedure 86
4.2.4. Data Collection and Analysis 87
4.3. Results 90
4.3.1. Behavioral data 90
4.3.2. Electrophysiological data 95
4.4. Discussion 96
4.5. Conclusion 101
5. General Discussion 103
5.1. Summary of the research studies 104
5.2. Limitations of current research 109
5.2.1. The measurement of emotional processes 109
5.2.2. Is research on moral dilemmas really impactful? 110
5.3. Mechanisms behind decisions and judgments in moral dilemmas 111
5.3.1. Anticipatory emotions 113
5.3.1. Anticipated emotions 114
5.3.2. Cost-benefit cognitive analysis 116
5.3.3. Representation of norms 117
5.3.4. The neural substrates of decision-making and judgments in moral dilemmas 119

5.4. Conclusions 123

6. References 125

1. GENERAL INTRODUCTION

Imagine you have to make a moral decision. If you had to sacrifice one person in order to save five, would you do it?

This is a question which has received great attention in psychology and neuroscience in the last fifteen years. When faced with this question, people generally think: "of course I'm going to sacrifice one person if I can save five lives, because that's the rational thing to do". But humans are not always rational beings, and they often decide in non-rational ways, based on preferences that originate from non-rational processes. One of these processes, and maybe the one that influences decisions in the most insidious way, is emotion. The research studies presented in this thesis are about what role emotions play in decision-making processes in the moral domain.

This introductory chapter aims at giving an overview of the literature from which the studies presented in this thesis took their basis: it encompasses the role of emotion in decision-making processes in general, the role of emotion in moral decisions and judgments, and which factors influence the emotional reactions to different morally relevant situations, with a particular focus on the relationship between emotion and intentionality of actions.

1.1. Emotion and decision-making

Intuitively, we feel that moral judgments have something special that distinguish them from judgments pertaining to other domains, like aesthetic judgments or probability judgments. Moral judgments just feel more important and universal to us. However, research on morality indicates that the mechanisms involved in moral decisions and judgments are not different from those involved in decisions and judgments in other domains (Young & Dungan, 2012). For instance, there are clear parallelisms between the valuation mechanisms implied in economic and moral decisions, and the brain systems involved in these two domains are largely overlapping (e.g., Hutcherson, Montaser-Kouhsari, Woodward, & Rangel, 2015). For this reason, I will start by describing some

theoretical models of the role played by emotion in decision-making that can be applied also to the moral context.

1.1.1. The somatic marker hypothesis

One of the most influential contribution on the role of emotion in decision-making is the somatic marker hypothesis (e.g., Bechara & Damasio, 2005), which posits that adaptive decision-making processes are not only based on a rational cost-benefit analysis, but also on the automatic and unconscious anticipation of emotional consequences. As we will see in the next paragraphs, the somatic marker hypothesis is particularly important for understanding decision-making in moral situations, since our moral decisions often seem to contradict a rational cost-benefit analysis and to follow an emotional evaluation instead.

The somatic marker hypothesis was formulated after observing the behavior of patients with ventromedial prefrontal cortex (vmPFC) lesions. These individuals, despite maintaining a normal IQ and normal problem-solving abilities in controlled settings, show remarkable impairments in judgment and decision-making in real life. For instance, they have difficulties in maintaining meaningful personal relationships and in pursuing advantageous behaviors in their private life and career, often incurring in severe financial and social losses and being apparently unable to learn from previous mistakes (Bechara, Damasio, Tranel, & Anderson, 1998; Damasio, Tranel, & Damasio, 1990; Eslinger & Damasio, 1985). The somatic marker hypothesis attributes these impairments to a deficit in an emotional mechanism. Normally, this mechanism supports decision-making by signaling the expected consequences of a choice through somatic markers.

Somatic markers may be elicited by two categories of situations, the first of which involves a type of stimuli, the so-called primary inducers, that are automatically (and often innately) able to elicit a somatic response. These stimuli can be both pleasurable (like some fatty and sugary food) or aversive (like a dangerous animal), and they elicit the activation of subcortical structures like the amygdala, which would then trigger the generation of a body state that allows a quick response to

the primary inducer (e.g., a fight/flight response in the case of a dangerous animal) and, consequently, a conscious emotional experience (e.g., fear). The second category of circumstances, the so-called secondary inducers, consist of memories and thoughts of primary inducers. According to the somatic marker hypothesis, remembering or imagining emotionally relevant situations causes the activation of the vmPFC, which, during lifetime, has associated emotionally relevant stimuli with the somatic states that they have elicited in the past. The vmPFC would then activate, through its connections with the amygdala, a body state alteration that would be a fainter version of the emotion related to the correspondent primary inducer. Through the vmPFC, for instance, imagining a loss of money re-activates the same pattern of somatic state that was experienced during an actual money loss experience. When we are faced with a choice (for instance between making or not making a hazardous gamble), this emotional mechanism helps us choose the best option by linking the available alternatives with somatic states that were associated in the past with positive or negative outcomes. The immediate and proximate consequences of an action would be able to directly activate the amygdala, but the future and more abstract consequences of a choice would only be able to trigger somatic states through the activation of the vmPFC. Thus, when pondering a complex decision in a context of ambiguity or uncertainty, patients with lesions in the vmPFC are biased toward the immediate consequences, because those would be the only consequences able to elicit a somatic state (Bechara, Damasio, Tranel, & Damasio, 2005).

It is worth mentioning that, according to the hypothesis, the somatic states are able to drive decision-making processes towards an option even before we consciously assess the consequences that would result from that choice. Without this mechanism, and in the absence of immediate consequences, the only possible way to make decisions for vmPFC patients would be a deliberate cost-benefit analysis. This would make it impossible for them to make advantageous choices in a reasonable time, especially in ambiguous and uncertain situations (Bechara & Damasio, 2005).

A series of empirical findings supports the somatic marker hypothesis, showing that individuals with vmPFC damage fail to present anticipatory SCRs when making risky decisions. In

these experiments, participants with vmPFC lesions were faced with a task that was specifically designed to mimic the kind of real world decisions in which these individuals show impairment – the Iowa Gambling Task (IGT). In this task, participants are asked to draw one hundred cards, choosing freely between four decks on which they are given no information. Two of the decks yield a high magnitude of gains and losses, and a negative net outcome; the other two yield a low magnitude of gains and losses, but a positive net outcome. Thus, the most advantageous strategy in the game would be to pick from the last two decks. In the studies by Bechara, Damasio, Damasio and Anderson (1994), healthy individuals learned to choose from the last two decks, and were successful in the task. Participants with vmPFC lesions, on the other hand, kept on drawing from the first two decks. Thus, vmPFC patients seemed more focused on the large immediate rewards that they could receive from the disadvantageous deck, and were unable to anticipate the future consequences of picking from those two decks.

Further work showed that healthy control participants exhibited anticipatory SCRs during the deliberation time before selecting a card, and that the magnitude of these SCRs was larger when participants subsequently selected a card from one of the two disadvantageous decks. Shortly after this difference in magnitude appeared, participants started to prefer the advantageous decks. This effect could be observed even if participants were unaware of their strategy. Patients with vmPFC damage, instead, did not show anticipatory SCRs for neither of the two decks, and never learned to pick from the advantageous decks. Thus, these findings support the role of the vmPFC in generating somatic markers that drive decisions towards advantageous choices. Without being able to mark the possible future outcomes with a positive or negative somatic state, individuals are not able to take future consequences into account when they are making a decision (Bechara et al., 1994; Bechara, 1997).

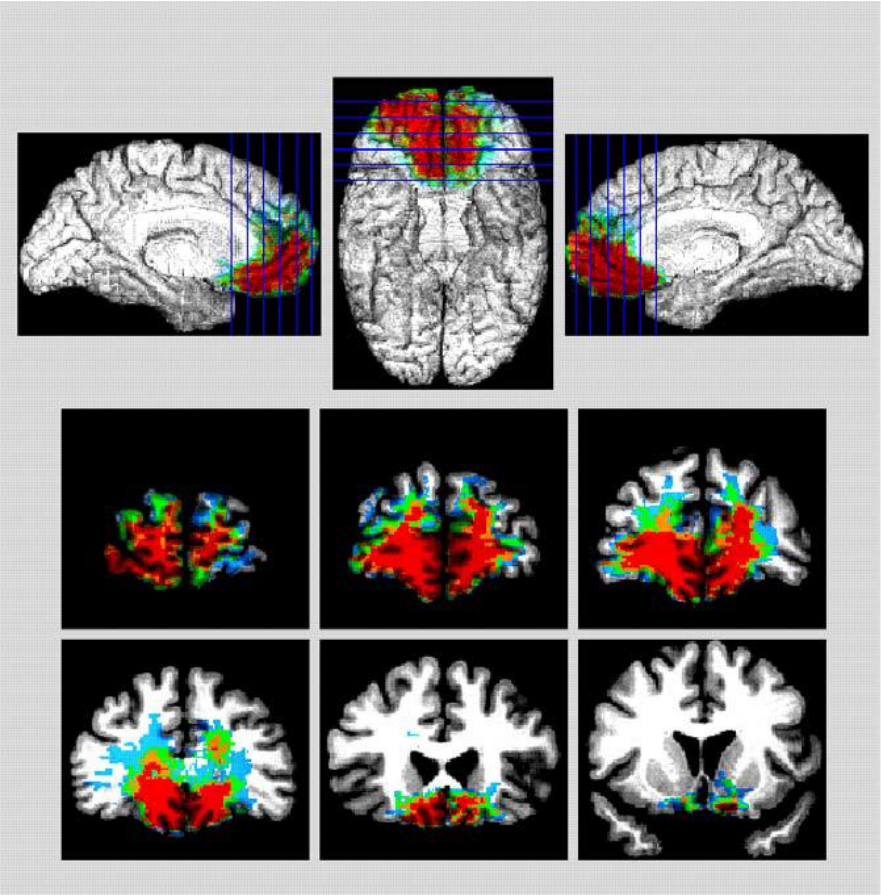


Figure 1.1. Overlap of vmPFC lesions of 13 patients who were administered the IGT. The red color indicates an overlap of four or more patients. Taken from Bechara & Damasio, 2005.

1.1.2. Decision-related emotions

Another line of evidence on the relevance of emotions in decision-making comes from the study of decision-related emotions, that is emotions which are specifically elicited in the context of a decision, like regret and disappointment (Zeelenberg & Pieters, 1999). Regret results from counterfactual comparisons between alternatives (that is, from the comparison of "what is" with "what might have been", Byrne, 2002; Epstude & Roese, 2008) and is elicited by thinking that the outcome of the chosen option is worse than what would have resulted from the alternative ones. Disappointment results from the comparison between the outcome achieved and an expected or desired one, and is elicited by the actual outcome being worse than the expected one. The function of these emotions is to modify future behavior based on previous experiences: before making a decision, people try to anticipate whether they would feel regret or disappointment as a

consequence of their choice, and use this information as input in the decision-making process (Bell, 1982, 1985; Coricelli & Rustichini, 2009; Loomes & Sugden, 1982, 1986). Regret seems to be more relevant than disappointment in modifying future behavior, since, as opposed to disappointment, it is strongly related to feelings of responsibility (Frijda, Kuipers, & ter Schure, 1989; Zeelenberg, van Dijk, & Manstead, 2000).

Emotions like regret and disappointment could be considered like the conscious and cognitively mediated equivalents of the somatic states proposed by the somatic marker hypothesis. Coherent with this interpretation, there is evidence that the vmPFC is a common substrate of the induction of somatic markers and of the experience of regret and of its anticipation: for instance, experiencing regret activates an area of the orbitofrontal cortex (OFC) that overlaps with the vmPFC (Coricelli et al., 2005), and patients with lesions in this area are unable to experience this emotion (Camille, 2004). According to Loewenstein and colleagues (Loewenstein & Lerner, 2003; Loewenstein, Weber, Hsee, & Welch, 2001) the difference between counterfactual emotions like regret and disappointment and somatic markers would be that the somatic markers are *anticipatory* emotions, that is, immediate visceral reactions that mark the future outcome of a decision, which are felt *at the moment* of the decision. Regret and disappointment, on the other hand, are *anticipated*, or *expected*, emotions, that is, emotions that people expect to feel *in the future*, but do not feel at the moment of the decision. However, anticipatory and anticipated emotions are strongly linked: for instance, anticipating the regret that we would feel if we made a risky choice could itself elicit immediate visceral reactions (Loewenstein & Lerner, 2003).

In the moral domain, emotions like guilt and shame seem to play a similar role as regret, since they are also elicited by a negative outcome and play a role in changing future behaviors (Wagner, N'Diaye, Ethofer, & Vuilleumier, 2011). Moreover, the emotion of guilt is influenced by similar factors as regret, as the sense of agency and personal responsibility, and shares some neural substrates like the OFC (Wagner et al., 2011) The main difference between guilt, shame and regret is that guilt and shame, but not regret, arise specifically in social contexts (Tangney, Miller, Flicker,

& Barlow, 1996; Wagner, Handke, Dörfel, & Walter, 2012). The function of guilt seems to be that of preventing people to incur in interpersonal harm and motivate people towards cooperative behavior (de Hooge, Zeelenberg, & Breugelmans, 2007; Wagner et al., 2012), while the function of shame seems to be to make people behave in compliance with social rules and social standards (de Hooge et al., 2007; Haidt, 2003). The role of these emotions in driving moral behavior will be described in more detail in Chapter 4.

1.2. Emotion and morality

Moral judgment has been long considered the result of a process of conscious, rational reasoning (Kohlberg, Levine, & Hewer, 1983; Vera-Estay, Dooley, & Beauchamp, 2014). In the last fifteen year, this view has been more and more challenged by the idea that morality is indissolubly tied to intuitions (e.g., Haidt, 2001) and to emotional processing (Blair, White, Meffert, & Hwang, 2013; Greene, 2008; Haidt & Bjorklund, 2008; Nichols, 2002; Prinz, 2006). The rest of this paragraph will focus on how emotion-based intuition, often in competition with deliberate thinking, influences moral judgement and moral decision. This paragraph also describes how the role of emotion in morality seems to be especially tied to harm aversion, that is, to the fact that the suffering of others is a naturally aversive stimulus for humans. Moreover, the role that the attribution of intentionality has in modulating moral judgments, moral decisions, and emotional reactions to harm will also be addressed.

1.2.1. Intuitive processes in moral judgment

A long tradition of research in psychology posits a dual process route to decisions and judgment: on the one hand, there is a rational and deliberate "System 2", that produces decisions and judgments through logic and rule-based reasoning processes (Sloman, 1996; Stanovich & West, 2002; Tversky & Kahneman, 1983); on the other hand, an intuitive, rapid and largely unconscious "System 1" that drives judgment and choices through innately programmed instinctive behaviors (Stanovich, 2004), heuristics (Kahneman & Frederick, 2002; Sloman, 1996; Stanovich & West,

2002; Tversky & Kahneman, 1983), and emotional processing (Epstein, 1994). People can usually coherently motivate only choices and judgments produced deliberately through System 2, whereas they can't provide satisfactory explanations when they try to motivate judgment and choices driven by System 1 (see Evans, 2003, 2008, for reviews). This seems to be especially relevant for the moral domain, as research in the last fifteen years has increasingly pointed out situations in which moral judgments and decisions don't seem to be supported by conscious deliberation, but, on the contrary, seem to be the result of rapid, and often emotionally laden, intuitions.

Some first results in these directions came from the studies of Haidt and colleagues. In a first pioneering study, Haidt, Koller and Dias (1993) presented to a sample of individuals from different socio-economical background a series of vignettes describing harmless, but in some way offensive, actions, such as cleaning a toilet with a national flag, eating one's own dead dog or masturbating with a dead chicken. Results showed that a majority of individuals judged those actions as morally wrong despite their harmless nature. Importantly, in a subsequent study, inducing disgust in participants through hypnosis before presenting them with the vignettes, made moral judgments more severe (Wheatley & Haidt, 2005). From these results, and taking inspiration from the somatic marker hypothesis, Haidt and colleagues (Haidt & Bjorklund, 2008; Haidt et al., 1993; Haidt, 2001; Wheatley & Haidt, 2005) proposed the *social intuitionist model*, according to which moral judgment are almost exclusively the product of emotional reactions, and rational deliberation has only the function of providing a post-hoc justification for the judgment. According to the model, reasoning can influence moral judgment, but only very slowly through discussions and confrontations with social peers that would first modify our emotional evaluation of some actions, and secondly, as a consequence, the moral judgment (see Figure 1.2). Haidt and colleagues acknowledge that reasoning may have a direct influence on our emotionally-laden intuitions, but this would occur only very rarely (Haidt & Bjorklund, 2008; Haidt, 2001).

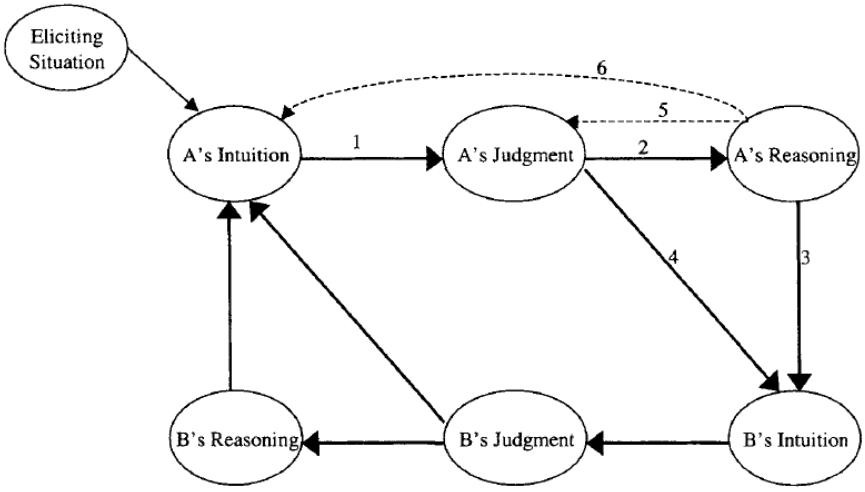


Figure 1.2. The social intuitionist model of moral judgment. The numbered links, drawn for Person A only, are (1) the intuitive judgment link, (2) the post hoc reasoning link, (3) the reasoned persuasion link, and (4) the social persuasion link. Two additional links are hypothesized to occur less frequently: (5) the reasoned judgment link and (6) the private reflection link. Taken from Haidt, 2001.

The social intuitionist model gave rise to a huge amount of literature with the aim of investigating how emotions can influence moral judgment, and if it is true that rational reasoning doesn't play a relevant role in it. An alternative approach that will be described in the next paragraph takes inspiration from dual process models and proposes a dual process model of moral judgment (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), in which emotional intuitions can drive decisions, but only when they survive a competition with slow and deliberate reasoning.

1.2.2. Moral dilemmas and the dual process model of moral judgment

Perhaps the biggest contribution on the importance of emotions in morality came from research on moral dilemmas. In these scenarios, originally devised by philosophers as an instrument to reason on different types of moralities (e.g., Foot, 1983; Thomson, 1985), individuals are asked to choose between two alternative options. Both of the options imply negative consequences, and none of them can be indisputably deemed as the morally right one (Braunack-Mayer & Joy, 2001; Sinnott-Armstrong, 1987).

A prototypical moral dilemma is the *Trolley problem*, in which a runaway railway trolley is about to run over a group of five unaware workers. The only way to save the five is to pull a lever and divert the trolley onto another rail, where a single worker stands, who would be run over instead (see Figure 1.3). A variant of the trolley problem is the *Footbridge problem*, in which the only way to save the five is to push a large stranger off an overpass, so that his body would stop the trolley (see Figure 1.3).

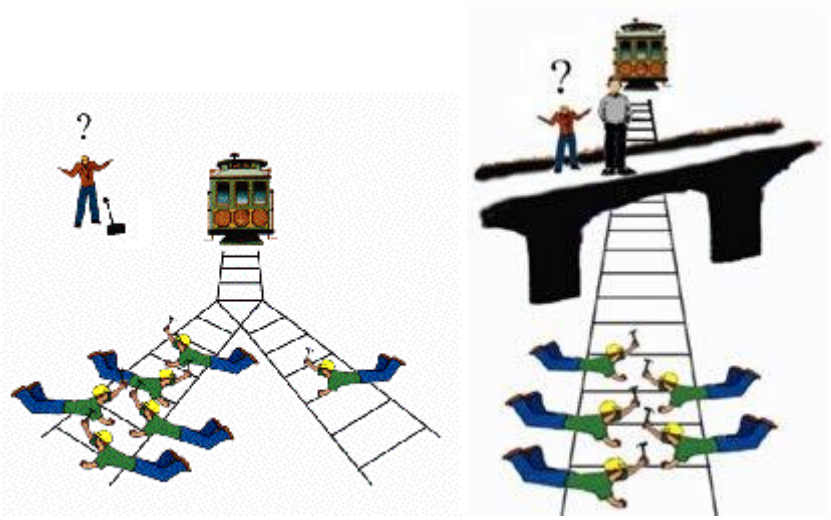


Figure 1.3. Graphical representation of the Trolley problem (on the left) and the Footbridge problem (on the right)

It is a now well-known and widely replicated result that individuals usually endorse the choice to sacrifice one person to save five lives in the Trolley dilemma, but not in the Footbridge dilemmas (Greene et al., 2001; Hauser, Cushman, Young, Kang-Xing Jin, & Mikhail, 2007; Sarlo et al., 2012; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). People are usually unable to adequately justify their choices: when they are asked why they endorse the choice of pulling the lever in the Trolley dilemma, people usually say that it was because saving as many lives as possible is the right thing to do. When they are asked why they didn't come to the same conclusion in the Footbridge dilemma, most of the people simply state that "it was a gut feeling", or that "killing is wrong" without being able to explain the incongruity, which suggests a role of intuitions in shaping this response pattern (Cushman, Young, & Hauser, 2006; Hauser et al., 2007;

Mikhail, 2002). Thus, in the Trolley dilemma people seem to prefer a "utilitarian" resolution, choosing the option that maximizes benefits and reduces costs. In the Footbridge dilemma, on the other hand, only very few individuals choose this same resolution. Thus, the question arises as to why people are utilitarian in the Trolley but not in the Footbridge dilemma.

A possible answer comes from the studies of Greene and colleagues (Greene et al., 2004, 2001). In these studies, participants underwent functional magnetic resonance imaging (fMRI) while responding to a set of sixty moral and non-moral dilemmas. The moral dilemmas were divided in three categories according to whether they were more similar to the Trolley dilemma or to the Footbridge dilemma (the criteria of the categorization will be described and discussed in detail in paragraph 1.2.3.). Dilemmas were presented in text form and participants could read at their own pace. For each dilemma, they had to judge if the action described (e.g., pushing a man off a bridge in the Footbridge dilemma, or pulling the switch in the Trolley dilemma) was morally "appropriate" or "inappropriate". The initial hypothesis of Greene and colleagues was that emotional processing would be more engaged by moral dilemmas similar to the Footbridge than by moral dilemmas similar to the Trolley, and this difference in emotional processing would affect people's judgments. The results were in line with the hypothesis: first of all, the percentage of "appropriate" judgments was higher for Trolley-type than for Footbridge-type dilemmas. Second, the analysis of response times showed that responses of "appropriate" in Footbridge-type dilemmas took longer than responses of "inappropriate", but there was no difference in response times between the two judgments in Trolley-type dilemmas. This result was interpreted by the authors as a consequence of the fact that judging the action proposed in Footbridge-type dilemmas as "appropriate" requires to overcome and control an automatic emotional response that would lead to judge the action as "inappropriate". Finally, the fMRI data were coherent with the hypothesis of a stronger emotional engagement in Footbridge-type as opposed to Trolley-type dilemmas: cortical areas including the vmPFC, the posterior cingulate cortex (PCC), and the superior temporal sulcus (STS), all related to emotional processing (Kosslyn et al., 1996; Maddock, 1999; Reiman, 1997;

Reiman et al., 1997) showed greater activation for Footbridge-type as compared to Trolley-type dilemmas. On the other hand, the dorsolateral prefrontal cortex (dlPFC) and the parietal lobe, areas related to abstract reasoning and working memory (Cohen et al., 1997; Smith & Jonides, 1997), were more active during Trolley-type than during Footbridge-type dilemmas (see Figure 1.4).

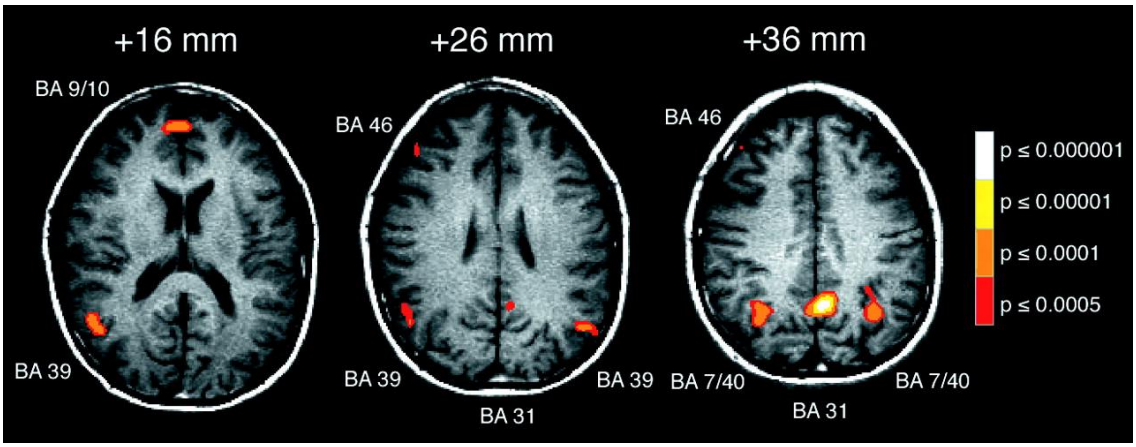


Figure 1.4. Brain areas exhibiting differences in activity between Footbridge-type and Trolley-type. Slice location is indicated by Talairach coordinate. Colored areas reflect the thresholded F scores. Images are reversed left to right to follow radiologic convention. Taken from Greene et al., 2001.

Based on this evidence, the authors proposed the *dual process model of moral judgment*, according to which moral judgment and moral decisions are the result of a competition between two processes: a slow, deliberate and rational process, that would perform a cost-benefit analysis and lead individuals to endorse the option that maximizes benefits and reduces costs, thus endorsing an utilitarian resolution of the dilemmas, and a fast, automatic and emotional process, that would work like a sort of "alarm bell" producing an immediate negative reaction against the proposed action (i.e., killing a man), leading individuals to reject the utilitarian resolution and to endorse the non-utilitarian resolution of the dilemmas.

It is important to remark that Greene and colleagues based their hypothesis that emotional processing plays a causal role on moral judgment entirely on brain activity. However, the cortical areas highlighted in these studies (the vmPFC, the PCC and the STS) are active during emotional processing task, but also during tasks involving other functions (i.e., they have high sensitivity for

emotional processing, but a low specificity). The vmPFC, in particular, is involved in a variety of other processes like self-representation (Ćurčić-Blake, van der Meer, Pijnenborg, David, & Aleman, 2015; Sui, Enock, Ralph, & Humphreys, 2015), schematic memory (Spalding, Jones, Duff, Tranel, & Warren, 2015) and internally guided decision-making (Nakao, Ohira, & Northoff, 2012), just to name a few. Moreover, recent findings seem to indicate that the vmPFC is more generally involved in integrating value information rising from different appraisals (not only of the strictly emotional type) in complex decisions (Grabenhorst & Rolls, 2011; Hutcherson et al., 2015; Rangel & Hare, 2010). For this reason, a reverse inference in this context is particularly risky. Thus, the studies of Greene and colleagues provided preliminary information that is coherent with an account proposing a role of emotion in moral judgment, but the evidence yielded by their studies is by itself not conclusive.

1.2.2.1. Studies testing the dual process model of moral judgment

In the last decade, a vast amount of research has been devolved in testing the dual process model of moral judgment, yielding generally coherent results indicating that emotions play a role in moral judgment and that they often are in competition with cognitive processes. However, the picture emerging from these studies seems to be more complex than how Greene and colleagues imagined it when they performed their first studies.

Evidence in line with the dual process model has been collected by several studies First of all, manipulating emotional and cognitive processing during the resolution of moral dilemmas influenced decisions and judgments in line with what predicted by the dual process model. For instance, inducing a happy mood in participants before a moral dilemma task produced more utilitarian judgments (Valdesolo & DeSteno, 2006). Moreover, individuals who performed a moral dilemma task under a cognitive load condition (i.e., performing a concurrent digit-search task) reported increased response times for utilitarian judgments as compared to individuals who performed a moral dilemma task with no concurrent task. Response times for non-utilitarian

judgments, on the other hand, were not influenced by the cognitive load manipulation. These results are coherent with the hypothesis that utilitarian judgments, but not non-utilitarian judgments, are driven by cognitive processes, since interfering with cognitive processes has a specific influence on utilitarian judgments only (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008).

Another line of evidence in support of the dual process model has been provided by studies that investigated the influence of individual differences in emotional and cognitive processing on utilitarian choices. For instance, the endorsement of utilitarian resolution in Footbridge-type dilemmas correlated positively with high trait psychopathy (Bartels & Pizarro, 2011; Gao & Tang, 2013; Glenn, Koleva, & Iyer, 2010; Koenigs, Kruepke, Zeier, & Newman, 2012; Patil, 2015; Seara-Cardoso, Dolberg, Neumann, Roiser, & Viding, 2013; Tassy, Deruelle, Mancini, Leistedt, & Wicker, 2013; Wiech et al., 2013, but see Cima, Tonnaer, & Hauser, 2010; Glenn, Raine, Schug, Young, & Hauser, 2009; Glenn, Raine, & Schug, 2009; Seara-Cardoso, Neumann, Roiser, McCrory, & Viding, 2012), a clinical construct characterized by emotional hyporeactivity and by a tendency to immoral behavior (Blair, 2013; Cleckley, 1976; Hare, 2003). Furthermore, the endorsement of utilitarian resolution in Footbridge-type dilemmas correlated negatively with empathic concern (Gleichgerrcht & Young, 2013), which is the tendency to experience feelings of sympathy and compassion in front of the suffering of others (Batson, 2009; Davis, 1983), and personal distress (Sarlo, Lotto, Rumiati, & Palomba, 2014), which is the tendency to experience discomfort and anxiety in front of the suffering of others (Batson, 2009; Davis, 1983). Also, high testosterone levels, which are associated with reduced harm aversion, were found to be positively correlated with utilitarian judgments (Carney & Mason, 2010). On the other hand, the number of utilitarian judgments was found to positively correlate with high working memory capacity and high need for cognition, a personality trait reflecting a preference for rational thought and effortful cognition (Moore, Clark, & Kane, 2008; Wiech et al., 2013). Taken together, these results are coherent with the predictions of the dual process model in suggesting that utilitarian judgments are

either due to a weak emotional reaction, or to a strong cognitive control that overcomes emotional processing.

Also, the investigation of the emotional state experienced during the resolution of the dilemmas yielded support for the dual process model. In a study by Sarlo and colleagues (2012) participants reported being in a more unpleasant state while they were deciding for Footbridge-type than for Trolley-type dilemmas. Coherent with this result, in a series of study by Szekely and Miu (2015), when participants decided to reject the utilitarian choice, they reported experiencing more intense negative emotions during the decision. Importantly, reducing emotional arousal through reappraisal strategies increased the number of utilitarian choices. Furthermore, Cushman, Gray, Gaffey and Mendes (2012) showed that individuals who present a higher autonomic reactivity during the resolution of the dilemmas, as indicated by increased total peripheral resistance, also present a lower endorsement of the utilitarian option.

Important evidence in line with the dual process model resulted also from an event-related potential (ERP) study by Sarlo and colleagues (2012), which provided information about the time course of the cognitive-emotional interplay of decision-making in moral dilemmas. In this study, ERPs were recorded time-locked to a decision slide during which participants were asked to choose between the utilitarian and the non-utilitarian resolution. The authors found an early positive component peaking at about 260 ms after the onset of this slide (P260), which was larger for Footbridge- than for Trolley-type dilemmas (Figure 1.5). Since the amplitude of this component positively correlated with the unpleasantness experienced by participants during decision-making, it was interpreted as a neural event signaling the strong negative emotional reaction that, according to Greene et al.'s (2004, 2001) dual process model, would drive decisions towards the non-utilitarian choice. In a subsequent time window, Trolley-type dilemmas were characterized by larger positive slow wave amplitudes as compared to Footbridge-type dilemmas, indicating greater allocation of attentional resources. Thus, this study provided additional support to the dual process model of moral decision (Greene et al., 2004, 2001) by showing how emotional processes "kicked-in" early

during decision-making, as indicated by the P260, and were followed by more controlled cognitive processes in a later stage, as indicated by the slow wave amplitudes. Moreover, a high disposition to react with personal distress to the suffering of others was associated with a larger the P260 amplitude and with a higher percentage of non-utilitarian choices, specifically in Footbridge-type dilemmas (Sarlo et al., 2014). This last result suggests that the "alarm-bell" emotional reaction might signal the anticipation of the personal distress that would result from choosing the utilitarian option.

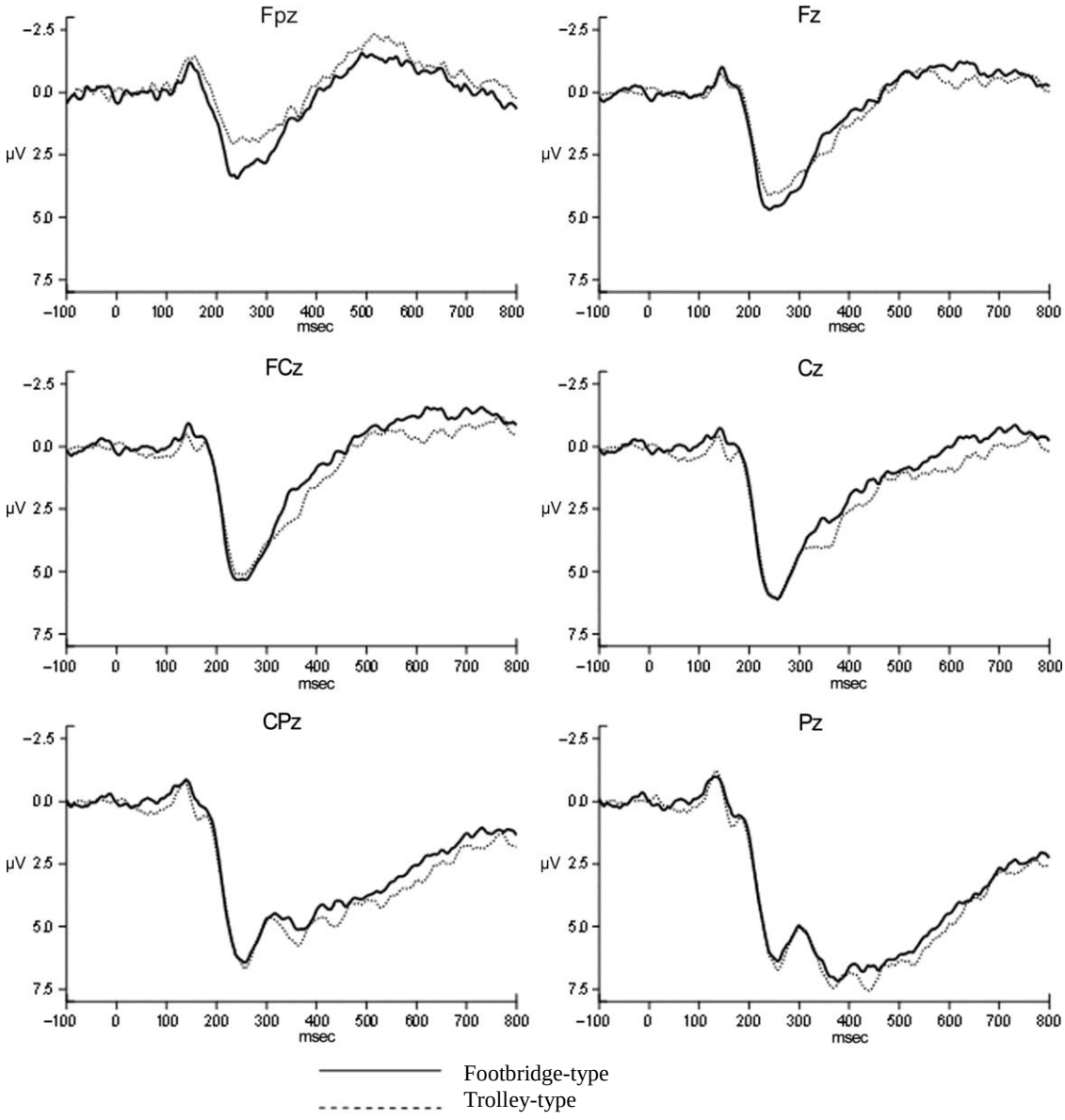


Figure 1.5. Grand-averaged ERPs recorded by Sarlo and colleagues (2012) at representative midline sites. Time 0 indicates the onset of the decision slide.

Finally, the dual process model found support in several neuropsychological studies. (Mendez, Anderson, & Shapira, 2005) administered the Trolley and the Footbridge moral dilemmas to a sample of patients with the frontal variant of fronto-temporal dementia. This neuropsychological syndrome is characterized by emotional blunting, lack of empathy and impaired social functioning, ranging from loss of social tact to sociopathic behavior, as a result of a progressive degeneration of the frontal lobes (Mendez, Chen, Shapira, & Miller, 2005; Neary et al., 1998). The judgments that these patients provided when faced with the Trolley and the Footbridge moral dilemmas were compared with those of a healthy control group and those of a group of patients with Alzheimer disorder, who showed a comparable level of cognitive impairment, but did not present specific emotional deficits. As compared to both groups, patients with fronto-temporal dementia were more utilitarian in the Footbridge dilemma, but not in the Trolley dilemma. Since early fronto-temporal dementia especially affects the vmPFC, this result shows that the integrity of the vmPFC is necessary for the rejection of the utilitarian option in the Footbridge dilemmas (Mendez, Anderson, et al., 2005). As seen in paragraph 1.1.1, according to the somatic marker hypothesis the vmPFC has a crucial function in decision-making since it triggers automatic aversion signals that allow individuals to avoid potentially dangerous choices. Without the correct functioning of the vmPFC, the future emotional consequences of the choices are not taken into account in the decision-making process (Bechara & Damasio, 2005). Thus, the results reported by (Mendez, Anderson, et al., 2005) are in line with what hypothesized by the dual process model of moral judgment.

Further evidence in this direction comes from studies by Koenigs and colleagues (2007) and Moretto, Làdavas, Mattioli and di Pellegrino (2010), who found that patients with vmPFC lesions were more utilitarian in Footbridge-type dilemmas than control participants. Moreover, Moretto and colleagues (2010), by recording SCRs during the dilemma task, showed that participants with vmPFC lesions failed to show anticipatory SCRs when solving the dilemmas. According to the somatic marker hypothesis (Bechara & Damasio, 2005), in a decision-making context SCRs are

triggered by the vmPFC and operate like alarm signals, marking the options associated with future negative consequences so that they can be avoided. Thus, like the patients studied by Bechara and colleagues (2005) were unable to avoid choosing risky options because they failed to present anticipatory SCRs to those options, participants in the study of Moretto and colleagues (2010) seemed to be unable to reject the utilitarian options in Footbridge-type dilemmas, because they failed to present anticipatory SCRs to them.

Similar results were found for patients with addiction to alcohol, who were more utilitarian than controls. These results have been interpreted as due to the fact that long-term alcohol exposure damages the vmPFC and, as a consequence, causes emotional dysfunctions (Khemiri, Guterstam, Franck, & Jayaram-Lindström, 2012). Carmona-Perera, Reyes del Paso, Pérez-García and Verdejo-García (2013) also administered a set of moral dilemmas to patients with alcohol addiction, and measured heart rate changes during the task. Their results are in line with those obtained by Khemiri and colleagues (2012),as participants with alcohol addiction were more utilitarian than control participants in Footbridge-type dilemmas, but not in Trolley-type dilemmas. Moreover, as opposed to controls, they did not discriminate between Footbridge-type and Trolley-type dilemmas in terms of heart rate responses.

However, not all the research studies that tested the dual process model provided results in line with its predictions. Other studies showed that the role of emotions in moral decisions might be more complex than what initially hypothesized by Greene and colleagues (2004, 2001). For instance, a study by Ugazio, Lamm and Singer (2012) showed that the influence of emotion on moral judgment depends on the motivational implication of the experienced emotion: for both Trolley- and Footbridge-type dilemmas, inducing anger (an approach-related emotion entailing a motivation to act) increased the endorsement of the utilitarian options, whereas inducing disgust (a withdrawal-related emotion entailing a motivation to retreat) reduced the endorsement of the utilitarian options. Similar results were reported by Choe and Min (2011), who found that trait disgust and trait empathy correlated negatively with utilitarian moral judgments, whereas trait anger

correlated positively. Moreover, a study by Moore, Stevens and Conway (2011) reported that participants with high behavioral approach sensitivity were more utilitarian in moral dilemmas, and that participants with high behavioral inhibition sensitivity were less utilitarian. According to the authors, these results reflect the fact that a high behavioral approach sensitivity is related to high sensitivity to gains and rewards, whereas a high behavioral inhibition sensitivity is related to high sensitivity to losses and punishments. Thus, individuals who are highly sensitive to gains evaluate saving more lives as more acceptable even when the tradeoff implies killing one person. Individuals who are highly sensitive to negative information, on the other hand, evaluate saving more lives as acceptable only when killing is less aversive (i.e., in Trolley-, but not in Footbridge-type dilemmas).

Other contrasting results came from the study of Patil, Cogoni, Zangrando, Chittaro and Silani (2014), who found that presenting Trolley-type dilemmas as virtual reality scenarios, rather than as written vignettes, increased emotional arousal, but also utilitarian choices. This is in contrast with what the dual process model would hypothesize, since according to the model heightened emotional arousal would result in a lower number of utilitarian choices. Moreover, in a study by Terbeck and colleagues (2013) reducing autonomic arousal through beta-adrenergic blockade reduced utilitarian judgments in Footbridge-type dilemmas instead of increasing them, as the dual process model would predict. Thus, the relationship between the arousal experienced during the resolution of the dilemmas is not straightforward, since it was reported to be associated both with an increase in utilitarianism – for instance, in the study by Moretto and colleagues (2010) previously described – and a decrease in it – like in these last two studies.

Finally, recent fMRI studies by Shenhav and Greene (2014) and by Hutcherson and colleagues (2015) provided interesting information suggesting a revision of the neural substrates of the dual process model. In these studies, brain activation associated with the emotional evaluation of the dilemmas, with the utilitarian evaluation, and with all-things-considered moral judgments, was investigated separately. In the study by Shenhav and Greene (2014), participants were presented with moral dilemmas and, in different trials, were asked to rate on four-points scales

either which options they felt worse about doing (emotional assessment), which option they thought would produce better results (utilitarian assessment) or which option they found more morally acceptable (moral judgment). In the study by Hutcherson and colleagues (2015), in the emotional and utilitarian assessment blocks, participants were presented with single options separately (e.g., either "Push large man in front of a runaway bus, killing him instantly", or "Prevent injury of five pedestrians by a runaway bus") and for each one they had to rate either how appealing/appalling it was (emotional assessment) or how costly/beneficial it was (utilitarian assessment). Subsequently, in the moral judgment block, they were presented with the two options depicted one against the other (e.g., "Push large man in front of a runaway bus, killing him instantly" in order to "Prevent injury of five pedestrians by a runaway bus") and they had to rate how these two deeds together were morally inappropriate/appropriate. Results showed that the emotional and utilitarian value of each option was computed in parallel in distinct brain areas, and subsequently integrated into an overall moral judgment in the vmPFC. The insula, the anterior cingulate cortex (ACC), the superior temporal gyrus (STS) (Hutcherson et al., 2015) and the amygdala (Shenhav & Greene, 2014) were found to be involved in emotional appraisal, whereas the temporoparietal junction (TPJ) and the dorsomedial prefrontal cortex (dmPFC) were found to be involved in the utilitarian appraisal (Hutcherson et al., 2015). This shows that the emotional and the utilitarian appraisals of the dilemmas are indeed encoded in two independent neural systems. The information that they separately encode provides weights in favor of each of the two options, which are integrated in the vmPFC, so that the option with the highest overall weight is selected. These results suggest a revision of the initial formulation of the model, which suggested that the vmPFC was one of the areas involved in the emotional evaluation (Greene et al., 2001). Rather, they are in line with a recent view of the vmPFC that depicts this cortical area as responsible for integrating goal-relevant information into a common value signals that guides decision choices (Grabenhorst & Rolls, 2011; Rangel & Hare, 2010). This view requires an alternative explanation of the fact that patients with vmPFC damage are more utilitarian than control participants. A possibility that has been

hypothesized by Shenhav and Greene (2014) is that deliberate reasoning can influence decision-making through an alternative route that does not include the vmPFC, whereas emotional appraisal needs to be integrated in the vmPFC to be able to exert an influence in judgments and choices. However, it is not clear which this alternative route would be, and this hypothesis has still to be tested.

1.2.3. What makes a moral dilemma a Footbridge-type dilemma? Dilemmas characteristics and categorization

The studies presented up to now showed how Trolley-type and Footbridge-type dilemmas differentially engage emotional processing and give rise to different judgment and choices. But which psychological features of the Trolley and the Footbridge dilemmas are able to produce these differences? What makes a dilemma a "Footbridge-type" dilemma or a "Trolley-type" one?

A first categorization was proposed by Greene and colleagues (2001), who divided their set of dilemmas in "personal" and "impersonal" according to three criteria schematically described as "ME HURT YOU". First of all, "personal" dilemmas must entail physical harm (the "HURT" criteria), that occurs to someone that is vividly represented as an individual (the "YOU" criterion). Finally, this harm must be the result of diverting an existing threat onto a different party, but has to be "authored" and not "edited" by the agent (the "ME" criterion). "Impersonal" dilemmas just don't fulfill one or more of these criteria. For instance, the Trolley dilemma doesn't fulfill the "ME" criterion, because the single worker's death would be only a side effect of having diverted the trolley onto another track. An impersonal dilemma used by Greene and colleagues (2001) that doesn't fulfill the "YOU" criterion is, for instance, the vaccination dilemma, in which the agent decides to make obligatory for the population to undergo a vaccination that will save the majority of people from a dangerous illness, but will kill a certain amount of people, which are represented just as numbers and not vividly as individuals. In the stimulus set used by Greene and colleagues (2001) there are also impersonal dilemmas that don't fulfill the "HURT" criterion, like the "taxes"

dilemma, in which the protagonist decides to evade taxes and doesn't physically harm anyone. The problem with this categorization is that there is significant heterogeneity both between the "personal" and "impersonal" categories and within them, especially within the "impersonal" one, so it's unclear which of the different features of the scenarios influenced the subject's response, and how: is it the "HURT" criterion, the "ME", the "YOU" or all of them combined that elicit the "alarm bell" emotional response?

Greene and colleagues (2001) proposed the personal/impersonal distinction to make a preliminary step toward the identification of the elements that influence moral judgment through the engagement of emotional processing. However, most of the studies on moral dilemmas kept using the personal and impersonal distinction and the stimulus set created by Greene and colleagues (2001) instead of attempting to identify a more precise categorization. Only a minority of studies focused on investigating which psychological features of the dilemmas were able to produce a difference in choices, which brain areas were involved in the processing of these features and if participants were aware of those differences between dilemmas and were able to attribute their choices to them. Three main factors were hypothesized as able to influence moral judgments and decisions in moral dilemmas: the physical contact with the victim, the personal force with which the victim is sacrificed, and the means vs side effect distinction.

1.2.3.1. Physical contact and personal force

One of the most evident differences between the Trolley and the Footbridge dilemma is that the Footbridge dilemma requires physical contact between the agent and the victim, whereas the Trolley dilemma does not. For this reason, a few studies focused on the role of physical contact in producing the different pattern on judgments that can be observed in these dilemmas. A first study by Cushman and colleagues (2006) found that physical contact with the victim indeed influenced moral judgment, so that sacrificing a person through physical contact was judged as les acceptable than sacrificing a person without physical contact. However, a more recent study by Greene,

Cushman and Stewart (2009) indicates that this result might not be due to simple physical contact, but to a less obvious variable: personal force, that is, whether or not the sacrifice is performed through applying physical force to the victim, for instance through pushing with hands or with an instrument. This study systematically investigated the effect of three different factors on moral judgments in moral dilemmas: physical contact, spatial proximity with the person that is sacrificed, and personal force. Interestingly, spatial proximity and physical contact were not sufficient to produce a significant difference in judgments when differences in personal force were accounted for. On the other hand, irrespective of physical contact and spatial proximity, sacrificing a man by applying personal force was judged as less acceptable than sacrificing a man without personal force.

Importantly, the effect of personal force was modulated by a further parameter, that is, whether the sacrifice of one person was the necessary means to save more lives or a foreseen but undesired side effect: when the sacrifice was a means, personal force decreased the moral acceptability of the action, whereas when the sacrifice was a side effect, personal force produced no differences in judgments.

1.2.3.2. Means vs side effect

As already mentioned in the previous paragraph, one subtle, but relevant, distinction between the Trolley and the Footbridge dilemmas is the fact that in the Footbridge dilemma the death of one person is a means, intentionally used to save five people, whereas in the Trolley dilemma it is a side effect of saving five people. Research that focused on this principle found that the majority of individuals judged as not permissible to kill a man as a means to save more lives, but as permissible to kill a man as a side effect of saving more lives (Greene et al., 2009; Hauser et al., 2007; Mikhail, 2002). Moreover, individuals are generally unaware of the means vs side effect distinction and are unable to justify their judgment referring to this principle (Hauser et al., 2007; Mikhail, 2002). This shows that the principle according to which we judge harm as a means as worse than harm as a side effect is not a product of conscious reasoning, but rather an intuition.

Importantly, the effect of the means vs side effect distinction on moral judgment was found to be significant also after controlling for potentially confounding variables, like proximity and physical contact between the agent and the person to be sacrificed (Hauser et al., 2007). The influence of the means vs side effect distinction on judgments, choices and emotional processing was investigated also by Lotto, Manfrinati and Sarlo (2014), who devised a set of 60 moral dilemmas categorized as Trolley-type or Footbridge-type based on the means vs side effect distinction.

Schaich Borg and colleagues (2006) investigated which brain areas are involved in processing the means vs side effect distinction in sacrificial moral dilemmas. In coherence with the results reported by Greene and colleagues for personal and impersonal moral dilemmas (Greene et al., 2004, 2001), harm as a means to an end elicited more activity in emotion-related areas like the vmPFC and the STS, and less activity in areas associated with cognition like the parietal lobe.

The means vs side effect distinction effectively modulates judgments even when people consider non-physical harm, like violations of other people's property or rights. For instance, when considering objects owned by other people, individuals judge as less acceptable to destroy a single object to save five objects when it is a means than when it is a side effect. Expectedly, when the objects are not someone's property (and thus when no harm is done to anyone by destroying them), sacrificing one object to save five objects is never judged as unacceptable (Millar, Turri, & Friedman, 2014).

The main reason why the means vs side effect distinction affects moral judgment is that it affects the attribution of intentionality: in a study by Cushman and Young (2011) killing a person as a means to an end was judged by participants as being more intentional than killing a person as a side effect. The means vs side effect distinction affected attributions of intentionality also in non-moral vignettes, in which nobody was harmed. The fact that the means vs side effect distinction influences the attribution of intentionality makes this distinction particularly interesting and relevant, since intentionality is a crucial aspect for moral decisions and judgment, as we will see in paragraph 1.2.4.

1.2.4. The role of harm aversion and intentionality in causing emotional reactions in the moral domain

In the previous paragraph I discussed some of the features of Trolley-type and Footbridge-type dilemmas that are able to elicit differences in moral judgments and decisions. However, according to the dual process model of moral judgment (Greene et al., 2004, 2001), those features would not influence judgment directly, but rather through the stronger emotional reactions that they elicit. In this paragraph we will see why those features can be expected to produce differences in emotional reactions.

First of all, it is important to consider that avoiding harm to others is one of the most relevant moral obligations and seems to be universally endorsed by humans (see, for instance, Turiel, 2002). According to some authors, this would be the result of harm aversion, that is, the phenomenon by which humans find unpleasant to cause or even witness distress to other people. For instance, according to the Integrated Emotion System theory (IES, Blair, 1995, 2007a), emotional aversion to harm is necessary for the development of morality. According to the IES, individuals learn to discriminate between good and bad actions due to the fact that other people's emotional expressions are intrinsic reinforcers or punishers. Expressions of fear or sadness, for instance, are intrinsic punishers. Through reinforcement learning, the actions that cause harm to others are represented as "bad". For instance, if a child pushes another child, making it cry, the crying of the latter child produces a negative emotional reaction in the former. This negative emotional reaction becomes associated with the act of pushing someone, that is thereby represented as "bad" and avoided.

Based on Blair's theories, recent research (Cushman et al., 2012; Cushman, 2013; Miller, Hannikainen, & Cushman, 2014) propose to distinguish between aversion to harmful *outcomes*, that is, to the suffering of others itself, and aversion to harmful *actions*, that is, to performing a harmful action in first-person. Moreover, some evidence points toward the presence of aversion to harmful

intent. As we will see, physical proximity, personal force and means vs side effect can modulate emotional reactions to the dilemmas through these three factors.

1.2.4.1. Aversion to harmful outcomes

A conspecific in distress is a relevant stimulus that naturally produces aversive reactions, not only in humans, but also in other mammals (e.g., Masserman, Wechkin, & Terris, 1964; Rice & Gainer, 1962; Sagi & Hoffman, 1976). This phenomenon is considered one of the main causes of helping and prosocial behavior. In humans, the mechanisms through which aversion to others' distress causes prosocial behavior are at least two: on the one hand, others' suffering elicits compassion and sympathy; on the other hand, witnessing others' suffering produces feeling of personal distress (Batson, 2009). Both of these feelings would motivate us to comfort the person who is suffering and to remove the cause of her suffering (Preston & de Waal, 2001). According to the IES (Blair, 1995, 2007a), the aversion to harmful outcomes is the necessary prerequisite that allows us to mark actions as "good" or "bad". However, as authors like Nichols (2002) pointed out, the link between harm aversion and judgment of harm as immoral is not as straightforward as it may seems: harmful outcomes are not the only features allowing us to judge a behavior as morally permissible or impermissible. For instance, if a person is harmed by accident, we might judge the event as "bad", but not as morally impermissible. Thus, aversion to harmful outcomes by itself can explain why we refrain from performing harmful actions, but not how we produce moral judgements.

Going back to moral dilemmas, both Footbridge- and Trolley-type dilemmas entail harmful outcomes, irrespective of the choice taken. However, the harmful outcome of sacrificing a person in Footbridge-type dilemmas could generate stronger emotional reactions than in Trolley-type dilemmas: imagining to kill someone as a means to an end, in close physical proximity and using personal force, possibly generates a more vivid imagery of the victim than imagining to kill someone who is physically distant, as a side effect, and without applying personal force. This more

vivid representation could increase empathy (see Singer & Lamm, 2009, for a review on the mechanisms of empathy), thus making the outcome emotionally more aversive. Importantly, physical proximity, personal force, and the means vs side effect distinction can all be hypothesized to independently increase the visual imagery of the victim. The same mechanism would also make killing a person more aversive than letting five people die in Footbridge-type dilemmas. In Trolley-type dilemmas, on the other hand, the visual imagery of the person to be sacrifice and that of the people to be saved would be equally vivid, and thus equally aversive. In line with this account, a study by Amit and Greene (2012) found that non-utilitarian judgments are supported by visual imagery of the victim to be sacrificed and that interfering with imagery through visual interferences increases utilitarian judgments.

1.2.4.2. Aversion to harmful actions

According to Cushman and colleagues (2012), the aversion to others' distress *per se* is not sufficient to explain some of the findings of research on morality. These authors state that the aversion to harm does not only consist of aversion to seeing or imagining other people suffering, but also of aversion to performing harmful actions *per se*, which would be acquired, as suggested by Blair (e.g., Blair, 1995, 2007b), through associational learning between some particular actions and the aversive stimulus represented by others' distress. In line with this theory, individuals who showed aversion to harmful actions even in the absence of harmful outcomes (e.g., participants who showed stress responses when firing a toy gun at the experimenter) were also less utilitarian in sacrificial moral dilemmas (Cushman et al., 2012). Moreover, the effect of high trait psychopathy in increasing the number of utilitarian responses to sacrificial moral dilemmas seems to be mediated by action aversion, which is inversely correlated with trait psychopathy (Patil, 2015).

Action aversion could be one of the factors accounting for the strong effect of the means vs side effect distinction on judgments and decisions in sacrificial moral dilemmas, since sacrificing a person as a means generally requires performing a prototypically harmful action (e.g., pushing the

man in the Footbridge dilemma), whereas sacrificing a person as a side effect does not (e.g., pulling the lever in the Trolley dilemma). Moreover, the concept of action aversion is linked to both physical proximity and personal force, since it is likely that most of the prototypically harmful actions imply both of these factors.

However, it is still unclear which features of an action would acquire an association with harm and consequently elicit aversion. According to Cushman and colleagues (Cushman et al., 2012; Miller & Cushman, 2013), the basic perceptual and motor properties of an action play a crucial role. In this case, however, it would be hard to explain why people acquire aversion to firing a toy gun, since it is unlikely that they had the opportunity to associate the motoric properties of such an action with harm in their life. At the same time, it would be hard to explain why people feel aversion to kicking a person, but not to kicking a ball, since the motoric properties of those actions are the same. In any case, the results presented by Cushman and colleagues (Cushman et al., 2012; Miller & Cushman, 2013) show that people are indeed more averse to perform a prototypically harmful action in first person than to witness the same action being performed by someone else, a result that can't be explained by outcome aversion alone. One alternative explanation of this effect could be that performing an action, instead of merely witnessing it, involves a sense of agency and personal responsibility that generate anticipated feelings of guilt and regret (Frijda et al., 1989; Wagner et al., 2012; Zeelenberg, van Dijk, et al., 2000). Moreover, the aversion to perform some actions in first person might be due to the fact that some actions, like pushing a man or firing a gun, are more strongly associated with harmful intents than others, like pulling a switch, are.

1.2.4.3. Aversion to harmful intents

A third and important element that could influence emotional reactions to moral dilemmas is intentionality. As we will see in this paragraph, several studies reported that the intentionality of an action is able to influence not only the moral judgment of that action and the attribution of blame

and punishment to the actor, but also the emotional evaluation of the outcome and of the actor himself.

First of all, a harmful action performed with the intention to harm is judged as morally worse than an action with the same harmful action that was performed unintentionally (Cushman, 2008; Ohtsubo, 2007). Coherently, intentionally harmful actions are assigned more punishment than accidentally harmful actions (Cushman, 2008). For instance, accidentally burning a partner's hand while trying to weld together two pieces of metal is judged as more morally permissible and less blameworthy than intentionally burning a partner's hand in the same context (Cushman, 2008). Moreover, bad intent *per se* is sufficient to judge an action as blameworthy, even when the planned harmful outcome doesn't occur (Young & Saxe, 2009b). For instance, attempting to put poison in a friend's coffee is judged to be blameworthy even when the attempt fails (Young & Saxe, 2009b).

Inferring the agent's intentions seems to be an intrinsic part of moral judgment. In fact, there is evidence indicating that moral judgment spontaneously elicits mental state inferences: Young and Saxe (2009a) found that when participants read vignettes describing morally relevant facts, brain areas involved in mental state inference, like the right TPJ, the precuneus, and the medial prefrontal cortex (mPFC), were more active as compared to when participants read vignettes describing morally irrelevant facts. The activity in these areas, the TPJ in particular, modulates moral judgment: using multi-voxel pattern analysis (MVPA), Koster-Hale, Saxe, Dungan and Young (2013) found that intentional and unintentional harms elicited differential patterns of activation in the right TPJ, and that individual differences in these patterns predicted individual differences in moral judgments so that the more the right TPJ discriminated between intentional and unintentional harm, the more the moral judgments were influenced by intentionality.

According to the results reported by Treadway and colleagues (2014), the TPJ is part of a network involving also the dorsal ACC, that suppresses the activity of the amygdala in response to harmful outcomes: in their study, in cases of intentional harm, increasing the emotional responses to harmful outcomes (by providing participants with emotionally graphic descriptions of the harm)

increased amygdala activity and strengthened connectivity between this structure and an area of the dorsolateral prefrontal cortex (dlPFC) involved in the attribution of punishment (Buckholtz et al., 2008; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Accordingly, this also increased the severity of the punishment determined by participants. Conversely, when the harm was unintentional, activity in the amygdala was suppressed by the dACC-TPJ circuit, and the emotional salience of the outcome had no effect on punishment.

Interestingly, the intentionality of an action influences also the evaluation of both its outcome and its agent. For instance, the experienced intensity of pain is modulated by the intentionality with which the pain is inflicted: in a study by Gray and Wegner (2008), participants evaluated an electric shock as more painful when they thought it was delivered intentionally by a confederate than when they thought it was delivered unintentionally. Moreover, in a study by Liljeholm, Dunne and O'Doherty (2014), participants rated the confederates who administered them an aversive liquid as less likable when they thought they were doing it on purpose than when they thought they were doing it unintentionally, and reported experiencing more anger as a consequence of the administration when they thought it was intentional than when they thought it was unintentional.

A study by Young and colleagues (2010) suggests that the vmPFC plays a relevant role in producing an adequate emotional reaction to harmful intents. In this study, patients with vmPFC lesions were able to provide moral judgments of intentional harm, non-harm and accidental harm that were comparable to those of control participants, but showed abnormal judgments of attempted harm, as they rated it as more morally permissible as compared to control participants. According to the authors, this could reflect the fact that patients with vmPFC lesions, as opposed to healthy participants, may not experience an aversive emotional reaction when faced with an agent's intention to cause harm. This would be due to their inability to trigger somatic markers in response to abstract concepts (Bechara & Damasio, 2005). The authors exclude the possibility that their results could be the consequence of a deficit in assessing intentionality, since vmPFC patients didn't

show any difference from controls in the judgment of accidental harm: even if the outcome was equally harmful, they judged intended harms as less appropriate than accidental harm. The results of this study suggest that the activation of the vmPFC, reported by Greene and colleagues (2004, 2001) for Footbridge-type dilemmas, might be due to the emotional evaluation of harmful intentions, rather than (or in addition to) emotional evaluation of harmful outcomes.

The influence of intentionality on emotional reactions to harm and on moral judgment is especially useful to understand why dilemmas like the Footbridge problem, in which harm to others is a means to an end, and thus perceived as more intentional (Cushman & Young, 2011), elicit stronger emotional reactions and more severe judgments of moral unacceptability than dilemmas like the Trolley problem, in which harm is a side effect, and thus perceived as less intentional (Cushman & Young, 2011). Moreover, even if there are no studies directly investigating this issue, it is plausible that personal force has also an influence on attributions of intentionality, since it is unlikely that harm as a result of applying muscular force to the victim could be perceived as unintentional.

1.3. Aims of the research project and outline of the studies

The previous paragraphs described data and models indicating that moral decisions and judgments, like decisions and judgments in other domains, are not the exclusive product of rational deliberation, but can be also driven by intuition. Several authors propose that these intuitions are emotionally-laden (e.g., Greene et al., 2001; Haidt, 2001). For instance, according to the dual process model of moral judgment (Greene et al., 2004, 2001), judgments (and decisions) are the result of a competition between an intuitive emotional system and a deliberate rational one. This model found support in studies on sacrificial moral dilemmas, in which a person is sacrificed in order to save a higher number of lives. According to the dual process model, when this sacrifice is particularly aversive, the emotional system would kick in and push towards a non-utilitarian choice (i.e., rejecting the idea of killing the man while letting a higher number of people die). Otherwise,

the rational system would prevail and push towards an utilitarian choice (i.e., saving the people while sacrificing the man, because this choice would limit the costs in terms of lives).

As seen in the previous paragraphs, sacrificing a person as a means to an end (as in the Footbridge dilemma, in which the agent is required to push a man off a bridge, stopping a runaway trolley with his body and thus saving five workers that would be run over) is judged as less morally acceptable than sacrificing a person as a side effect (as in the Trolley dilemma, in which the agent is required to pull a switch to divert the runaway trolley on a side track where it would kill only one worker) (Hauser et al., 2007; Lotto et al., 2014; Schaich Borg et al., 2006). The reason behind this effect is that sacrificing a person as a means to an end is perceived as more intentional than sacrificing a person as a side effect (Cushman & Young, 2011). This, in turn, can be hypothesized to generate:

- a stronger impact of the negative outcome (i.e., of the death of one person), since intentionality influences the evaluation of the outcome of an action (Gray & Wegner, 2008);
- a more negative evaluation of the action itself, since intentionally harmful actions are considered as deserving more blame and punishment than unintentionally harmful actions (Cushman, 2008);
- stronger anticipated feelings of guilt and regret, since these two emotions are greatly enhanced by the perception of agency and responsibility (Frijda et al., 1989; Wagner et al., 2012; Zeelenberg, van Dijk, et al., 2000).

Up to now, a vast amount of literature has yielded evidence in line with the dual process model of moral judgment (see previous paragraphs), but this evidence is not completely conclusive, and further investigation is required to test whether and how emotions drive moral judgments and decisions, and if differences in emotional processing are necessary conditions to produce the different pattern of decisions and judgments that have been reported for moral dilemmas that vary on the means vs side effect distinction.

As concerns the first point, that is whether and how emotions drive moral judgments and decisions, the literature is coherent with an account that posits a role of emotion in these processes. First of all, influencing the mood and the emotions of the participants influences moral judgments: for instance, inducing a positive mood before the moral dilemma task makes participants more utilitarian (Valdesolo & DeSteno, 2006), as does instructing participants to down-regulate their emotions during the task (Szekely & Miu, 2015). Secondly, participants with emotional hyporeactivity, like participants with vmPFC lesions and participants with high psychopathy traits, are more utilitarian (e.g. Koenigs et al., 2007, 2012; Moretto et al., 2010). Finally, brain areas associated with emotional processing, like the vmPFC, the STS and the amygdala, are more active in Footbridge-type than in Trolley-type moral dilemmas (Greene et al., 2004, 2001).

This evidence, however, is not enough to conclusively confirm that emotions play a causal role in moral judgment and decisions, as the dual process model hypothesizes. For one thing, the fact that modulating participants' mood during the task influences their judgments does not prove that the emotions intrinsically caused by the dilemma affect the judgment, because it only indicates that incidental affect experienced during the resolution of the dilemmas influences judgment. Moreover, only very few studies collected participants' emotional evaluations during the task (Choe & Min, 2011; Lotto et al., 2014; Sarlo et al., 2012; Szekely & Miu, 2015). Among these, only two studies (Lotto et al., 2014; Sarlo et al., 2012) tested if Trolley-type and Footbridge-type dilemmas elicited different emotional evaluations, with only one (Sarlo et al., 2012) finding higher unpleasantness evaluations for Footbridge-type than for Trolley-type dilemmas. Most importantly, none of these studies reported a significant association between emotional ratings and judgments or choices. Some studies found that autonomic or self-reported arousal was associated with a decrease in utilitarian responses (Moretto et al., 2010; Szekely & Miu, 2015), but others found opposite results (Patil et al., 2014; Terbek et al., 2013).

Finally, the fact that brain areas related with emotional processing are more active during the resolution of Footbridge-type dilemmas is not necessarily an indication of the fact that emotional

processing plays a role in the dilemma resolution. First and foremost, the vmPFC, the brain area most consistently found active during the resolution of moral dilemmas (Glenn, Raine, & Schug, 2009; Greene et al., 2004, 2001; Hutcherson et al., 2015; Schaich Borg et al., 2006; Shenhav & Greene, 2014) has a low specificity for emotional processing, being involved in several other processes (Ćurčić-Blake et al., 2015; Nakao et al., 2012; Spalding et al., 2015; Sui et al., 2015). Secondly, even if the activation of the vmPFC during the resolution of moral dilemmas did actually reflects emotional processing, this result *per se* might be not sufficient to indicate that emotional processing plays a causal role on the decision instead of being merely a consequence of it, or an epiphenomenon.

As concerns the second point, that is whether differences in emotional processing are necessary conditions to produce the different pattern of decisions and judgments reported for Footbridge- and Trolley-type dilemmas, the situation is ambiguous: as seen in paragraph 1.2.3, sacrificing a person as a means to an end is perceived to be more intentional than sacrificing a person as a side effect (Cushman & Young, 2011). As seen in paragraph 1.2.4, intentionality influences moral judgment and the attribution of blame and punishment (Cushman, 2008; Koster-Hale et al., 2013; Young & Saxe, 2009b). Moreover, intentionality influences the emotional evaluation of the outcome of an action (Gray & Wegner, 2008) and the influence of the emotional impact of an outcome on punishment (Treadway et al., 2014). Thus, the question arises as to whether the assessment of intentionality would be *per se* sufficient to cause a difference in decisions and judgments in moral dilemmas, when it does not elicit a stronger emotional reaction. On the one hand, studies on patients with vmPFC lesions seem to indicate that with no emotional reaction to harmful intents, intentionality does not affect judgment (Young et al., 2010). On the other hand, behavioral data collected with moral dilemmas show that participants with emotional hyporeactivity (i.e., participants with psychopathy, vmPFC impairment or alcohol addiction) still present a difference in judgment and choices between Footbridge-type and Trolley-type dilemmas, endorsing the utilitarian option in Trolley-type dilemmas more than in Footbridge-type dilemmas.

(Carmona-Perera, Clark, Young, Pérez-García, & Verdejo-García, 2014; Khemiri et al., 2012; Koenigs et al., 2012; Mendez, Anderson, et al., 2005, but see Moretto et al., 2010). If the difference in decisions and judgments in Footbridge-type dilemmas was only due to emotional processing, then these participants would show a reduction, if not an absence, of this effect, but this was not what the studies reported (with the exception of Moretto et al., 2010). There are two possible explanations for these findings: the first is that the emotional reactivity of these individuals is dampened, but not completely eliminated, and might still play a role in driving decisions and judgments; the second is that emotional processing is not the only factor that can lead to different conclusions in Trolley- and Footbridge-type dilemmas, because even without a strong emotional reaction, individuals might rely on information about differences in intentionality to judge the sacrifice of a man in Footbridge-type dilemmas as less acceptable than in Trolley-type dilemmas. Thus, it is still to be cleared how emotional processing interacts with intentionality appraisals in influencing judgments and decisions in moral dilemmas.

Finally, the fact that intentionality influences attribution of blame and punishment suggests that an additional factor could drive decisions and judgments in moral dilemmas, that is, the consideration of hypothetical legal consequences. Intentionally sacrificing a person, like in the Footbridge dilemma, could be considered as a sanctionable offense more than unintentionally killing a person as a side effect, like in the Trolley dilemma. Thus, participants' choices and judgments might be driven by this additional evaluation, that might play a role in causing the different judgments that are so consistently observed between Trolley- and Footbridge-type dilemmas. This could be especially true when the question is posed to participants from a first person perspective (e.g., "would you push a person off a bridge in order to save five workers?" or "is it morally appropriate for you to push a person off a bridge in order to save five workers?"). Up to now, no study systematically investigated if and how this variable influences the resolution of moral dilemmas and how it interacts with the emotional appraisal of the dilemma.

Starting from these bases, the main aim of the studies presented in this thesis was to investigate how the means vs. side effect distinction affects decisions and judgments in moral dilemmas through its effects on emotional processing, evaluations of legal consequences, and anticipated emotional consequences of the decision choice. Moreover, in two studies the neural activity associated with decision-making was investigated using both stimulus- and response-locked event-related potentials (ERPs). The high temporal resolution that characterizes ERPs allows to separately examine different stages of the decision-making process. In particular, stimulus-locked ERPs allow to focus on an earlier stage, in which the decision options are assessed and evaluated (cf. Ernst & Paulus, 2005; Paulus, 2005); response-locked ERPs, on the other hand, allow to focus on the last stage of the decision-making process, in which the final choice is selected and the corresponding action is prepared by the motor cortex (cf. Gluth, Rieskamp, & Büchel, 2013). As concerns stimulus-locked ERPs, we focused on the P260 component – already mentioned in paragraph 1.2.2.1 – that indicates an immediate emotional reaction during the earlier stages of decision-making, and on the late positive potential (LPP), that is related to allocation of attentional resources and working memory load (Rösler & Heil, 1991). This allowed us to investigate an earlier, possibly automatic, appraisal of the decision options and a subsequent and more controlled elaboration stage. As concerns response-locked ERPs, we focused in particular on the readiness potential (RP), an electrophysiological correlate of the preparation of voluntary movements (Shibasaki & Hallett, 2006). Previous studies indicated that, in moral situations, this potential may reflect a conflict during the last phase of decision-making. In particular, Sarlo and colleagues (2012) found a reduced amplitude of this potential, reflecting a reduced preparation to act, for Footbridge- than for Trolley-type dilemmas, a result that was interpreted as reflecting a higher conflict between choice options in the former dilemma category. Moreover, a study by Panasiti and colleagues (2014) reported a lower amplitude of this potential for spontaneous deception as compared to truth telling, a result that also points toward an interpretation of this potential as a correlate of conflict in a moral context.

The studies presented in this thesis employed a standardized set of moral dilemmas that was developed by Lotto and colleagues (2014), in which Footbridge-type and Trolley-type dilemmas are categorized according to the means vs side effect distinction. This dilemma set provides important advantages as compared to the one devised by Greene and colleagues (2001): as described in Paragraph 1.2.3, in the original dilemma set used by Greene and colleagues the dilemmas are roughly divided in Footbridge-type and Trolley-type dilemmas through three broad criteria that make it difficult to understand which specific features of the dilemmas are able to influence emotional processing, and through which mechanisms they affect choices and judgments. Moreover, in the original set, the two categories of dilemmas (personal/impersonal) differ in several potentially confounding variables, with personal dilemmas more often using an emotionally vivid language, more often resulting in the death of people and more often involving friends or relatives as characters than impersonal dilemmas. The set of stimuli devised by Lotto and colleagues, conversely, only includes dilemmas involving issues of killing and letting die, it never features friends or relatives as characters, and always uses a plain language. Finally, in this dilemma set other potential confounds like the length of the text of the dilemmas (measured in number of words and letters) and the number of victims are balanced in the two categories.

The aim of the first study presented in this thesis was to investigate how taking into account hypothetical legal consequences influences the decision-making process during the resolution of moral dilemmas, in particular as regards the choices, the emotional state experienced during the resolution of the dilemmas, and the neural correlates of decision-making. The neural correlates of decision-making were examined through both stimulus-locked and response-locked ERPs, which allowed to investigate which stage of the decision process was influenced by the legal evaluations: the appraisal of the options, or the final implementation of the action corresponding to the decision choice.

After clarifying how legal evaluations interact with the means vs. side effect distinction in influencing the decision-making process and the emotional impact of the dilemmas, the second

study aimed at investigating how trait psychopathy affects choices, judgment and the emotional state experienced during the decision in dilemmas that differ on the means vs side effect distinction. Psychopathy is a clinical construct characterized by both emotional hyporeactivity to harm and by a tendency to immoral behavior (Blair, 2011; Cleckley, 1976; Hare, 2003). Importantly, the immoral behavior that individuals with psychopathy present is hypothesized to be a consequence of their emotional hyporeactivity to harm (Blair et al., 2013; Blair, 1995, 2007b, 2011). Thus, in the second study, comparing participants with high trait psychopathy with participants with low trait psychopathy allowed to examine how a tendency to emotional hyporeactivity impacts the affective state elicited by the resolution of moral dilemmas, and the subsequent decisions and judgments.

Finally, to further clarify how emotional processing affects decision-making in moral dilemmas, the last study focused on the anticipated emotional consequences produced by to the two options (utilitarian and non-utilitarian) depicted in the dilemmas, and on their influence on choices. Additionally, in this last study time-locked ERPs were measured with the aim of further investigating the neural correlates of the last phase of decision-making, focusing on the Readiness Potential (RP). This was aimed at clarifying the functional significance of this potential in the context of morally relevant decisions, by testing if reflects a conflict between the anticipated emotional consequences of the two options depicted in moral dilemmas.

2. STUDY I: THE INFLUENCE OF LEGAL CONSEQUENCES ON NEURAL ACTIVITY, EMOTIONAL EXPERIENCE AND DECISION CHOICE IN MORAL DILEMMAS[1]

2.1. Introduction

As seen in the general introduction, moral dilemmas have been widely employed in cognitive and affective neuroscience to investigate how cognitive and emotional processes interact in producing moral judgments and choices (e.g., Greene et al., 2001; Greene, 2008; Koenigs et al., 2007; Schaich Borg et al., 2006). Several studies reported that most people endorse the choice of sacrificing one person to save five people in dilemmas like the Trolley dilemma, in which this sacrifice would be a side effect of saving five people, but not in the Footbridge dilemma, in which this sacrifice would be the means to save the five people (Greene et al., 2009; Hauser et al., 2007; Lotto et al., 2014; Mikhail, 2002; Sarlo et al., 2012; Schaich Borg et al., 2006). According to the dual process model of moral judgment (Greene et al., 2004, 2001), this response pattern is due to the fact that decision in moral dilemmas stems from a competition between a fast, automatic emotional system and a slower, deliberate cognitive system. According to the model, Footbridge-type dilemmas strongly involve emotional processes, while in Trolley-type dilemmas emotional processes are less strongly activated and cognitive processes prevail, driving a "rational", utilitarian choice that maximizes the number of spared lives.

Harm used as a means to an end, but not harm as a side effect, is supposed to be able to trigger this emotional response (Cushman et al., 2012; Cushman & Greene, 2012; Greene et al., 2009). However, harming someone as a means is also perceived by individuals as more intentional than harming someone as a side effect (Cushman & Young, 2011). Both in ordinary social life and in modern systems of criminal justice there is widespread agreement that, besides the severity of

[1] The material presented in this chapter has been partially published in Pletti C., Sarlo M., Palomba D., Rumiati R., Lotto L. (2015). Legal concerns affect neural activity and emotional experience during the resolution of moral dilemmas. Brain and Cognition 94, 24-31, doi: 10.1016/j.bandc.2015.01.004.

harm, intentionality strongly affects the process of assigning blame for a harmful action (e.g., Darley, 2009; Nadler & McDonnell, 2011). Specifically, more blame is assigned for a harmful behavior that was intentional than for a comparable one that was unintentional (Guglielmo, Monroe, & Malle, 2009; Ohtsubo, 2007; Shultz & Wells, 1985; Treadway et al., 2014). Moreover, affective responses to harm were found to increase punishment severity only when the harm was intentional (Treadway et al., 2014).

By differing for the degree of intentionality, Footbridge- and Trolley-type dilemmas would thus promote differential assignment of blame and punishment while implicating the same severity of harm. On these bases, it might be argued that, in addition to the powerful negative emotional reaction hypothesized by Greene et al. (2004, 2001), the need to avoid punishment contributes to the rejection of the utilitarian choice in Footbridge-type dilemmas. Surprisingly, empirical investigation on the legal implications of this type of moral dilemmas has been largely neglected.

It is thus still unclear whether and how legal concerns for the harm to be committed contribute to decision-making when people evaluate moral dilemmas, nor whether such evaluations involve rapid, intuitive processes rather than effortful cognitive control. Indeed, whereas in the legal processes of trial and conviction decision-making involves a complex integration of cognitive evaluations with representations of relevant legal codes (Buckholtz & Marois, 2012; Schleim, Spranger, Erk, & Walter, 2011), in everyday life the assignment of blame and punishment seems to be based more on intuitive and implicit processes than on reasoning (Darley, 2009).

In the present study, we compared the neural activity, subjective emotional reactions and behavioral choices in two groups of participants who either took (*Legal group*) or did not take (*No Legal group*) legal consequences into account when deciding on moral dilemmas. In particular, we were interested in comparing the temporal dynamics of moral decision-making in people who believed or not that utilitarian resolutions had consequences in terms of legal liability. To this aim, we used an experimental paradigm (Sarlo et al., 2012) allowing to measure event-related potentials (ERPs) during two distinct phases of decision-making: one in which the dilemma resolutions were

assessed and compared (i.e., stimulus-locked ERPs), and one in which the corresponding action was selected, prepared and executed (i.e., response-locked, movement-related potentials, MRPs). Among the MRP components, we were particularly concerned with exploring the amplitude of the Readiness Potential (RP) as a function of group and dilemmas. The RP is known to reflect the preparatory processes preceding the execution of voluntary movements (Shibasaki & Hallett, 2006), with several studies indicating that it shortly precedes the conscious intention to act, thus reflecting intentional actions at preconscious stages (Haggard & Eimer, 1999; Haggard, 2005; Libet, Gleason, Wright, & Pearl, 1983; Rigoni, Kühn, Sartori, & Brass, 2011; Sirigu et al., 2004). This sustained negative potential is suggested to arise from the medial frontal areas, particularly from the supplementary motor area (SMA) (Cunnington, Windischberger, Deecke, & Moser, 2003; Praamstra, Stegeman, Horstink, & Cools, 1996). Lastly, in order to measure participants' emotional reactions, affective valence and arousal experienced during decision-making were collected after each decision.

We hypothesized that the evaluation of legal consequences would drive behavioral choices toward non-utilitarian resolutions, especially in Footbridge-type dilemmas, since they entail intentional harm. If this would be the case, then the Legal group would show a stronger preference for the non-utilitarian choice in Footbridge-type dilemmas as compared to the No Legal group. Moreover, the Legal group would show a greater difference in choices between Footbridge-type and Trolley-type dilemmas as compared to the No Legal group.

Furthermore, by taking the legal consequences into account, the Legal group would have an additional element on which to base its decisions, and thus it would show a facilitation of the decision-making process, as reflected in greater action readiness and lower decision times as compared with the No Legal group. On the other hand, by excluding legal consequences from the decision-making process, we hypothesized participants in the No Legal group to rely mainly on the affective information when deciding on Footbridge-type dilemmas, thus showing greater emotional engagement than the Legal group, at both neural and subjective levels.

Finally, a critical question was whether evaluations related to legal concerns would be reflected in the early or late ERP components before responding, thus affecting more automatic versus controlled processing stages.

2.2. Methods

2.2.1. Participants

A total of thirty-four healthy participants (20 women) aged 19–28 years (*M*= 24.2, *SD* = 2.2) were included in the present study. Participants included in the Legal group (*N* = 17) were selected from a previous study using identical experimental stimuli, paradigm, and procedure (Sarlo et al., 2012), as having reported in a post-experimental debriefing questionnaire to have taken legal implications into account when deciding on moral dilemmas. Specifically, participants were selected according to the following criteria: (a) having answered ‘‘yes’’ to the question ‘‘When deciding, have you ever thought about the possible legal consequences that your choice would have implied? For instance, have you thought that killing one person to save others might have resulted in homicide conviction?’’; (b) having reported a score > 2 on a 0–7 Likert scale, with 0 = not at all and 7 = completely, in which they had to indicate how much this idea had influenced their choices in the task.

Participants assigned to the No Legal group (*N* = 17) were newly recruited to match the Legal group according to age, gender, and educational level. They were instructed that none of the options proposed in the dilemmas was legally prosecutable. At the end of the task, they completed the same debriefing questionnaire that had been administered to the Legal group. All participants in the No Legal group who answered ‘‘yes’’ to the question pertaining criterion (a) were excluded from the study and replaced (*N* = 7).

All participants were right-handed and had no history of psychiatric or neurological disorders. They were paid €13 for their participation. The study was approved by the local ethical committee, and was conducted in compliance with the declaration of Helsinki.

2.2.2. Stimuli

We employed a set of 60 standardized dilemmas (Lotto et al., 2014) comprising 30 Footbridge-type dilemmas, which described killing one individual as an intended means to save others, and 30 Trolley-type dilemmas, which described killing one individual as a foreseen but unintended consequence of saving others.

Dilemmas were displayed in written text in a series of three slides: the scenario, in which a threat endangers several people's lives; Option A (non-utilitarian choice), in which the agent lets these people die; Option B (utilitarian choice), in which the agent kills one person to save these people.

Twelve additional moral dilemmas, which involved no deaths and described other moral issues (e.g., stealing, lying, and being dishonest), were used as filler stimuli. The experimental task consisted of 3 blocks, each comprising 10 Footbridge-type, 10 Trolley- type, and 4 filler dilemmas, presented in a pseudo-random order on a 19 inch. monitor at a viewing distance of 100 cm. Stimulus presentation was accomplished with E-prime software (Psychology Software Tools Inc., 2012).

2.2.3. Procedure

Upon arrival at the laboratory, participants read and signed an informed consent form. After an elastic cap was applied for EEG recording, instructions for the task were given. Each trial started with the scenario slide, which remained on screen until participants pressed the space bar. Subsequently, Option A was presented for 4.5 s, followed by Option B for 6.5 s. Thereafter, participants were presented with the decision slide, showing the letters A and B vertically aligned to a fixation cross in the middle of the screen, which remained on screen until participants responded, for a maximum time of 10 seconds. Participants were instructed to decide between the two options by pressing one of two computer keys marked ‘‘A’’ and ‘‘B’’. They were also explicitly instructed

to wait for the onset of this slide before starting to evaluate the options. The decision slide remained on screen for one additional second after the response, to prevent the MRPs to be contaminated by offset-related potentials. After their response, participants rated how they felt while they were deciding, using the 1–9 point scales of valence and arousal of the Self-Assessment Manikin (Lang, Bradley, & Cuthbert, 2008).

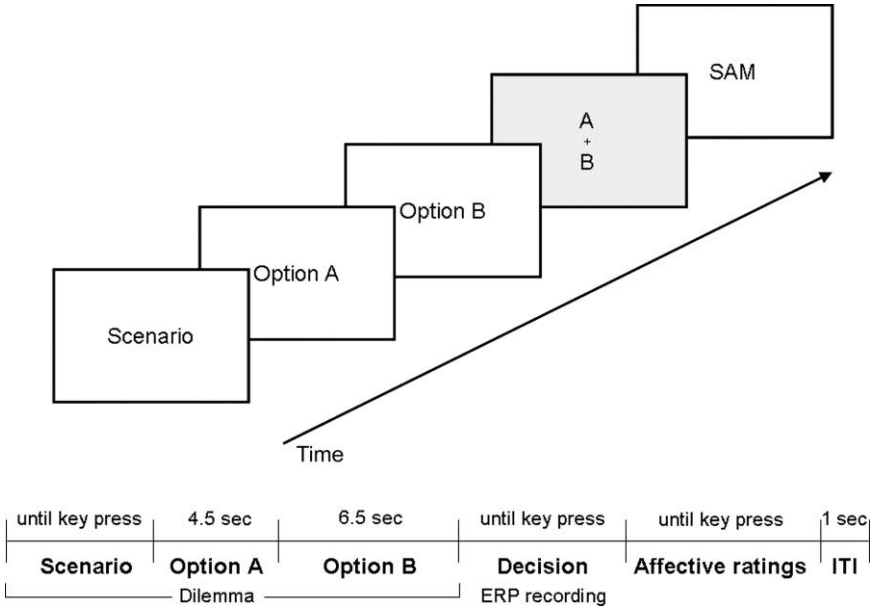


Figure 2.1. Sequence of events in the experiment. Participants had to decide between Options A and B by pressing the corresponding key during the presentation of the decision slide (in gray). ERPs were recorded time-locked to the decision slide onset. MRPs were recorded time-locked to the behavioral response. SAM = self-assessment manikin and ITI = intertrial interval. Text is not drawn to scale. Taken from Sarlo et al., 2012.

After two practice trials and before starting the experimental session, participants in the No Legal group were presented with a slide reporting the following statement: ‘‘According to the Article 54 of the Italian Penal Code[2], none of the options proposed in the dilemmas is legally prosecutable’’.

The task lasted about one hour, plus about forty minutes preparation time.

[2] Article 54 of the Italian Penal Code (1930/1987) provides: ‘‘No one shall be punished for acts committed under the constraint of necessity to preserve himself or others from the actual danger of a serious personal harm, which is not caused voluntarily nor otherwise inevitable, and the acts committed under which are proportionate to the threatened harm’’. Participants in the No Legal group, however, did not receive any information regarding the content of this Article until the end of the experimental session.

2.2.4. Data Collection and Analysis

The EEG was recorded from 31 tin electrodes (Fpz, Fz, FCz, Cz, CPz, Pz, Oz, Fp1, Fp2, F3, F4, FC3, FC4, C3, C4, CP3, CP4, P3, P4, O1, O2, F7, F8, FT7, FT8, T3, T4, TP7, TP8, T5, and T6) mounted in an elastic cap (Electro-Cap International, Inc.; Eaton, OH) and the right mastoid. All impedances were kept below 10 kΩ. All sites were referenced online to the left mastoid and re-referenced offline to the average of the left and right mastoids. For the purpose of artifact scoring, vertical and horizontal electro-oculograms (EOGs) were recorded. The EEG and EOG signals were amplified with Neuroscan Synamps (El Paso, TX), bandpass filtered (DC- 70 Hz) and digitized at 500 Hz (16 bit A/D converter, accuracy 0.1 μV per least significant bit). Blink artifacts and eye movements were corrected with a regression-based algorithm (Gratton, Coles, & Donchin, 1983). In order to compute ERPs, continuous EEG was segmented offline into 900-ms epochs from 100 ms before to 800 ms after the onset of the decision slide. To compute the MRPs, the EEG was segmented into 1500-ms epochs from 1000 ms before to 500 ms after the keypress. All epochs were linear detrended to correct for slow DC shifts, re-filtered with a 30 Hz low pass filter (12 dB/oct) and baseline-corrected against the mean-voltage recorded during the 100 ms pre-stimulus period for ERPs, and during a 200 ms period preceding keypress (from -1000 ms to -800 ms) for MRPs. The epochs were then inspected for artifacts, and every epoch containing a higher voltage than ±70 μV in any channel was rejected from further analysis. The remaining epochs were averaged separately for each participant and condition.

The P260 was identified and specified as the most positive peak between 200 and 300 ms from stimulus onset. Successive slow wave activity (i.e., the Late Positive Potential, LPP) was measured as mean amplitude within three successive post-stimulus time windows (LPP1, 300–450 ms; LPP2, 450–600 ms; LPP3, 600–750 ms). The amplitudes of MRP components were measured in three time intervals (cf. Shibasaki & Hallett, 2006): (1) mean negativity between 800 and 500 ms before key-press (early RP); (2) mean negativity between 500 and 50 ms before keypress (late RP); and (3) mean negativity between 50 ms preceding and 100 ms following key-press (motor potential,

MP). Separate ANOVAs were performed on mean LPP and MRP amplitudes, with Group (Legal, No Legal) as between-subjects factor and dilemma type (Footbridge-type, Trolley-type), Electrode Location (F, FC, C, CP, P) and Laterality (left, midline, right) as within-subjects factors. The ANOVA performed on mean P260 amplitudes and peak latencies focused on frontal locations (Fp, F, FC) based on previous results (Sarlo et al., 2012). Separate ANOVAs were conducted on the mean percentages of utilitarian choices, decision times, valence and arousal ratings, with Group as between-subjects factor and dilemma type as within-subjects factor.

The corrected p-values for effects within variables with more than two levels are reported together with the Huynh-Feldt epsilon (ε) and the uncorrected degrees of freedom. Bonferroni-corrected post hoc comparisons were conducted on significant main effects and interactions. Finally, Pearson's correlations between valence and percentage of utilitarian choices, and arousal and percentage of utilitarian choices were calculated separately by group and dilemma type. All statistical analysis were performed with IBM's SPSS (IBM Corp., 2010).

2.3. Results

2.3.1. Behavioral data

The ANOVA on the percentages of utilitarian choices revealed only a significant main effect of dilemma type ($F(1,32) = 168.23$, $p < .0001$, $\eta^2_p = .84$), with higher percentages for Trolley- than Footbridge-type dilemmas (*M*s = 75.04% and 38.31%, respectively). The ANOVA on decision times yielded no significant results (all *p*s > .43). Mean decision times were 2588 ms.

2.3.2. Affective ratings

The group main effect was significant for valence ratings ($F(1,32) = 4.79$, $p = .036$, $\eta^2_p = .13$) and marginally significant for arousal ratings ($F(1,32) = 3.83$, $p = .059$, $\eta^2_p = .11$), in that the No Legal group rated the decisions as more unpleasant and more arousing than the Legal group (Figure 2.2). The dilemma type effect showed a trend toward significance only for valence ratings

($F(1,32) = 3.24$, $p = .08$, $\eta^2_p = .09$), with decisions on Footbridge-type dilemmas rated as more unpleasant than those on Trolley-type dilemmas. None of the correlations between valence and percentage of utilitarian choice, and arousal and percentage of utilitarian choices was significant (all *p*s > .43).

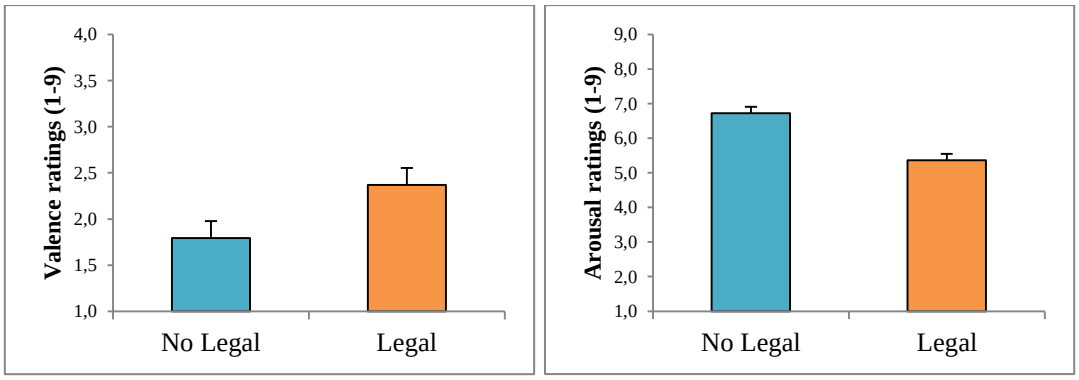


Figure 2.2. Mean valence (top) and arousal (bottom) ratings across dilemma types in the No Legal and Legal groups.

2.3.3. Stimulus-locked ERPs

Grand-averaged ERPs elicited during decision-making as a function of Footbridge- and Trolley-type dilemmas are displayed at representative midline sites in Figure 2.3.

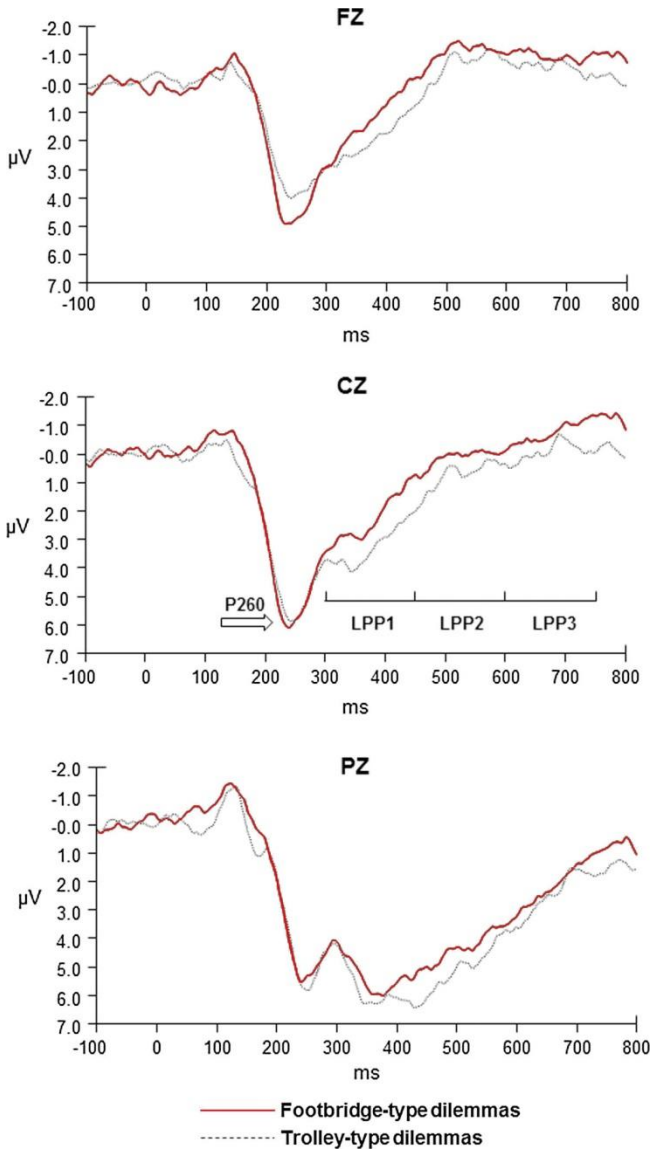


Figure 2.3. Grand-averaged ERPs recorded at representative midline sites time-locked to the decision slide for Footbridge- and Trolley-type dilemmas. Time 0 indicates the onset of the decision slide. The P260 component and the three time windows of the Late Positive Potential (LPP) are highlighted.

2.3.3.1. P260 amplitude and latency

No significant main effect or interactions involving the group factor were found for the P260 amplitude (all *p*s > .24). The significant main effect of dilemma type ($F(1,32) = 4.66$, $p = .039$, $\eta^2_p = .13$) showed that Footbridge-type dilemmas elicited larger amplitudes than Trolley-type dilemmas (*M*s = 6.07 and 5.39 μV, respectively; see Figure 2.3). The location main effect ($F(2,64) = 32.67$, $p < .0001$, $\varepsilon = .65$, $\eta^2_p = .51$) showed that the P260 was largest at fronto-central locations (*p*s < .0001). As further specified by the significant Location × Laterality interaction ($F(4,128) = 6.88$, $p < .0001$, $\varepsilon = .87$, $\eta^2_p = .18$), the P260 amplitude was maximal at FCz and FC4 (all *p*s < .005).

As for peak latency, no significant main effect or interactions involving the group factor were found. The significant laterality main effect ($F(2,64) = 8.65$, $p = .001$, $\varepsilon = .88$, $\eta^2_p = .21$) showed longer latencies on the right than on the midline and left sites (ps < .02).

2.3.3.2. LPP amplitudes

No significant main effect or interactions involving the group factor were found for the LPP1 amplitude (300–450 ms post-stimulus). The significant main effect of dilemma type ($F(1,32) = 5.1$, $p = .031$, $\eta^2_p = .14$) showed that Trolley-type dilemmas elicited larger amplitudes than Footbridge-type dilemmas (Ms = 3.49 and 2.86 μV, respectively; see Figure 2.3), indicating greater allocation of attentional resources. The location ($F(4,128) = 42.34$, $p < .0001$, $\varepsilon = .35$, $\eta^2_p = .57$) and laterality ($F(2,64) = 9.59$, $p < .0001$, $\varepsilon = 1.00$, $\eta^2_p = .23$) main effects were also significant. As further specified by the significant Location × Laterality interaction ($F(8,256) = 9.56$, $p < .0001$, $\varepsilon = .72$, $\eta^2_p = .23$), the LPP1 amplitude was maximal at Pz and P4 (all ps < .02). In the successive time windows (i.e., LPP2, 450–600 ms post-stimulus, and LPP3, 600–750 ms poststimulus), the only significant effects involved the location ($F(4,128) = 57.73$, $p < .0001$, $\varepsilon = .36$, $\eta^2_p = .64$; $F(4,128) = 16.48$, $p < .0001$, $\varepsilon = .37$, $\eta^2_p = .34$, respectively) and laterality ($F(2,64) = 11.44$, $p < .0001$, $\varepsilon = 1.00$, $\eta^2_p = .26$; $F(2,64) = 12.86$, $p < .0001$, $\varepsilon = .93$, $\eta^2_p = .29$) main effects, and the Location × Laterality interaction ($F(8,256) = 19.79$, $p < .0001$, $\varepsilon = .49$, $\eta^2_p = .38$; $F(8,256) = 21.20$, $p < .0001$, $\varepsilon = .44$, $\eta^2_p = .40$), indicating maximal amplitude at Pz and P4 for LPP2 and at Pz for LPP3 (all ps < .04).

2.3.4. Response-locked MRPs

Grand-averaged MRPs elicited before response choice in the Legal and No Legal groups are displayed at representative midline sites in Figure 2.4.

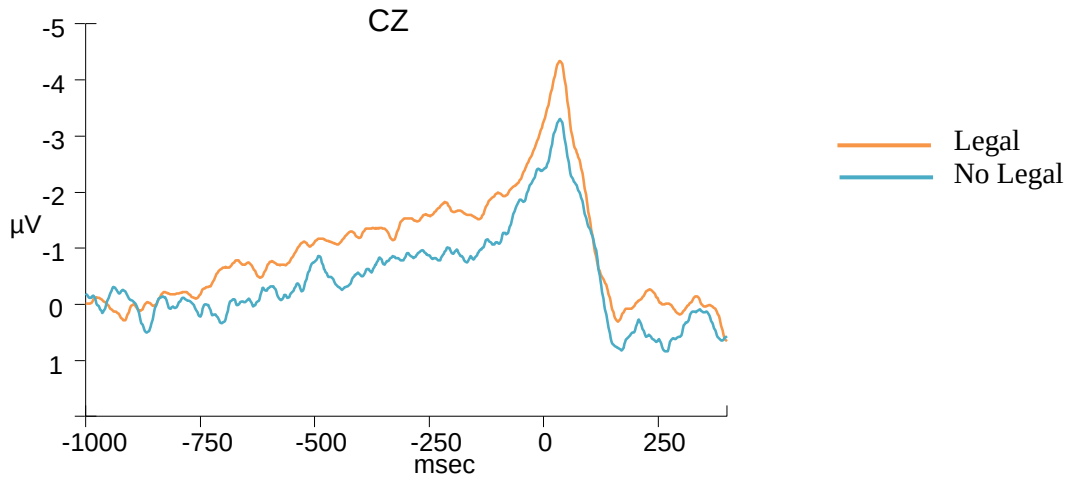


Figure 2.4. Grand-averaged MRPs recorded at representative midline sites time-locked to the behavioral response in the No Legal and Legal groups. Time 0 indicates the onset of the behavioral response.

2.3.4.1. Early RP

A significant group main effect emerged for the early RP amplitude ($F(1,32) = 5.43$, $p = .026$, $\eta^2_p = .15$), indicating greater negativity for the Legal than for the No Legal group (*M*s = -.44 and .21 μV, respectively) between 800 and 500 ms before key-press (see Figure 2.4). The significant Location × Laterality interaction ($F(8,256) = 3.14$, $p = .013$, $\varepsilon = .56$, $\eta^2_p = .09$) showed that the amplitude of the early RP was maximal at the midline central and centro-parietal locations (i.e., at Cz and CPz) (all *p*s < .05). A significant Group × Location interaction ($F(4,128) = 6.12$, $p = .012$, $\eta^2_p = .16$) was also found. Post-hoc comparisons between groups revealed that the Legal group showed larger amplitudes than the No Legal group at central, centro-parietal, and parietal locations (all *p*s < .03), whereas no differences were found at frontal or fronto-central areas. Post-hoc comparisons within groups revealed that the Legal group showed maximal amplitudes at the midline central and centro-parietal locations (all *p*s < .05), whereas the No Legal group did not show any significant difference in amplitudes among locations (all *p*s > .29).

2.3.4.2. Late RP

The group main effect was marginally significant ($F(1,32) = 4.0$, $p = .053$, $\eta^2_p = .11$), with greater negativity observed for the Legal than for the No Legal group (*M*s = -.82 and -.15 μV, respectively) between 500 and 50 ms before key-press. The significant Location × Laterality interaction ($F(8,256) = 4.42$, $p = .001$, $\varepsilon = .61$, $\eta^2_p = .12$) showed that the amplitude of the late RP was maximal at the midline central and centro-parietal locations (i.e., at Cz and CPz) (all *p*s < .04).

2.3.4.3. MP

The group main effect was not significant in the time window from 50 ms before to 100 ms after key-press. However, a significant Group x Location interaction ($F(4,128) = 7.10$, $p = .004$, $\eta^2_p = .18$) was found. Post-hoc comparisons between groups revealed that the Legal group showed larger amplitudes than the No Legal group at centro-parietal and parietal locations (*p*s < .04), whereas no differences were found at frontal, fronto-central, and central areas. Post-hoc comparisons within groups revealed that the Legal group showed maximal amplitudes at fronto-central and central locations (all *p*s < .04), whereas the No Legal group showed maximal amplitudes at frontal and fronto-central locations (all *p*s < .002). The significant Location × Laterality interaction ($F(8, 256) = 19.62$, $p < .0001$, $\varepsilon = .60$, $\eta^2_p = .38$) showed that the MP amplitude was maximal at the midline fronto-central and central locations (i.e., at Cz and CPz), and larger on the left than the right locations (all *p*s < .05).

2.4. Discussion

Despite the large number of studies that have systematically explored the critical factors affecting decision-making in Footbridge- and Trolley-type dilemmas (e.g., Bartels, 2008; Cushman et al., 2006; Greene et al., 2009; Mikhail, 2011; Moore et al., 2008; Schaich Borg et al., 2006), little attention has been devoted to the influence of evaluating the legal consequences the dilemma agent would face when choosing between resolutions. This is somewhat surprising, as in any modern society killing is regarded as a severe violation of the legal codes that is subjected to penal

judgment. On this basis, while this kind of moral dilemmas may be regarded as peculiar phenomena, because the violations are aimed at a greater good, it is implausible to assume that people do not take legal consequences into account when deciding about the hypothetical killing of one person. In particular, there is well- documented evidence that judgments of intentionality guide judgments of blame and affect punishment assignments in naïve individuals (e.g., Cushman, 2008; Darley & Pittman, 2003; Guglielmo et al., 2009), with emotional salience increasing the severity of punishment attributed to intentional, but not unintentional, harm-doing (Treadway et al., 2014). On these bases, we reasoned that Footbridge-type dilemmas, by implying greater intentionality than Trolley-type dilemmas (Greene et al., 2004, 2001; Sarlo et al., 2012; Schaich Borg et al., 2006), might evoke harsher blame and punishment for the proposed moral violations. Such evaluations might thus serve as further input to decision-making and add to Greene et al.'s (2004, 2001) ‘‘alarm-bell emotion’’ in driving behavioral choices toward the rejection of utilitarian resolutions.

The present study was aimed at comparing the temporal dynamics of moral decision-making in two groups of participants who reported being influenced in their decisions by the legal consequences of the proposed actions (Legal group) versus being instructed that none of the proposed actions had consequences in terms of legal liability (No Legal group). Decision-making was assessed across subjective, behavioral, and physiological domains, in order to highlight possible differences in explicit and implicit processing between groups. Moreover, the use of both stimulus- and response-locked ERPs allowed us to investigate the neural activity underlying two separate phases of the decision process: one associated with the initial evaluation and comparison of the dilemma’s options and one closely related to intention to act and response preparation.

A first relevant result to be noticed is that no difference in behavioral choices was found between the No Legal and the Legal group. Specifically, participants who were instructed that none of the options proposed in the dilemmas was legally prosecutable and participants who took legal implications into account chose a comparably lower proportion of utilitarian resolutions when deciding on Footbridge- than on Trolley-type dilemmas. Thus, legal consequences don’t seem to

play a confounding role in producing the difference in responses that is consistently observed between Trolley-type and Footbridge-type dilemmas (Greene et al., 2009; Hauser et al., 2007; Lotto et al., 2014; Mikhail, 2002; Sarlo et al., 2012; Schaich Borg et al., 2006).

However, this lack of a significant group difference in choices does not mean that legal consequences did not have an impact on the decision-making process: the dynamics of the psychological processes preceding the conscious motor decision appear to be different in the two groups, as supported by the relevant effects obtained for subjective and electrophysiological measures.

As hypothesized, at the subjective level, the No Legal group reported overall greater emotional engagement during decision- making relative to the Legal group, as revealed by feelings of higher unpleasantness and arousal. However, such a conscious emotional response was not reflected, at the neural level, in an overall larger amplitudes of the P260 for the No Legal than the Legal group. Consistent with previous research (Sarlo et al., 2012), both groups showed larger P260 and lower valence ratings when deciding on Footbridge- than on Trolley-type dilemmas (the latter effect being marginally significant). This ERP component might index the (aversive) affective appraisal of the choice options in the early processing stages, as its amplitude was found to positively correlate with the unpleasantness experienced during decision-making (Sarlo et al., 2012).

As indicated by the analysis of the LPP amplitudes, no difference between the No Legal and the Legal group was found even during the later processing stages associated with allocation of attentional resources and working memory load (e.g., Rösler & Heil, 1991). Consistent with previous research (Sarlo et al., 2012), both groups showed larger positivity during the resolution of Trolley- than Footbridge-type dilemmas, indicating that decisions required a larger amount of attentional resources.

Overall, then, no group differences emerged in the first phase of decision-making, as reflected in the amplitudes of the stimulus- locked ERP components up to 750 ms after the onset of

the decision slide. Specifically, taking the legal consequences into account during decision-making did not modulate the early processing stages nor did it require additional attentional resources in the later stages of option processing. In contrast, the effect of dilemma types, varying in the degree of intentionality and emotional engagement, largely prevailed in affecting neural activity, as well as behavioral choices. It seems plausible to suppose that this first phase of decision-making, which was unaffected by legal considerations but strongly modulated by the intentionality of actions, unconsciously laid the groundwork for selecting the dilemma's resolution. Subsequent evaluations possibly had effect on the regulation of conflict that is intrinsic to any resolution choice in this kind of moral dilemmas.

Indeed, legal concern exerted its effects in the last phase of decision-making just preceding the behavioral choice, in that the Legal group showed larger MRP amplitudes than the No Legal group, indicating greater readiness to act. In particular, the early RP component (until about 500 ms before movement onset) is an index of motor readiness modulated by the level of intention to perform a voluntary action (Shibasaki & Hallett, 2006) and was consistently found to shortly precede the conscious experience of intention to act (e.g., Haggard & Eimer, 1999; Haggard, 2005; Libet et al., 1983). Moreover, this electrophysiological response is known to originate primarily in the medial frontal SMA regions, which play a key role not only in motor planning, but also in higher cognitive processes, such as the anticipation of reward (Lee, 2004) or the encoding of values of available actions before choice (Wunderlich, Rangel, & O'Doherty, 2009). Importantly, and relevant to the present work, the amplitude of the early RP has been recently considered as a cortical marker of moral conflict in decision-making, as reduced motor preparation was observed in spontaneous deception as compared with truth-telling (Panasiti et al., 2014). In this view, the lower RP amplitudes observed in the No Legal than the Legal group might reflect reduced intention to act associated with greater (moral) conflict at the level of action preparation. Interestingly, this difference between groups was found to extend to the subsequent MRP components, thus affecting the whole process of motor preparation and execution. Indeed, larger amplitudes were found in the

Legal than the No Legal group both for the late RP and the MP components. The late RP (from 500 ms to movement onset) is known to indicate activity of the contralateral premotor and primary motor cortex and is influenced more by the implementation of specific features of movement, whereas the MP reflects the ultimate transmission of the descending motor command from the primary motor cortex (Shibasaki & Hallett, 2006).

Taken together, our results suggest that the behavioral choices were largely shaped across participants during the first phase of decision-making, probably being influenced by the emotional aversion to the act of intentionally sacrificing one person in order to save more lives. However, by excluding the evaluation of the legal consequences from the decision process, as conveyed by the experimental instructions, the No Legal group reported greater overall emotional impact, as compared with the Legal group. Interestingly, such emotional response pattern was associated with lower overall preparation for action, suggesting greater conflict between alternative motor responses representing the different decision choices. We might speculate that, with no reference to the legal consequences of the proposed violations, decision-making became more complex, thus increasing overall the emotional burden of the choice to make, at least at conscious level. In contrast, participants in the Legal group showed an overall dampened affective experience during decision-making associated with greater action readiness and intention to act, reflecting lower conflict in responding. On these bases, we might speculate that in moral dilemmas legal consequences of actions provide a sort of reference point on which people can rely to support a decision, independent of dilemma type. In contrast to what we hypothesized, however, no real facilitation of decision-making was found in the Legal group in terms of speeded decision, as no difference in response times were observed between groups. It is possible that comparable overall decision times in the two groups reflect different processing times in different phases of decision-making, which our paradigm was not sensitive enough to highlight. Indeed, despite investigating the initial and the last phase of decision-making, corresponding to the first 750 ms and the last 800 ms, we might have missed what is in between, as mean decision times were about 2600 ms.

As a main limitation of the present study, it should be noted that the two groups were not randomly determined, but rather were self-selected based on what they reported in the debriefing questionnaire, particularly the Legal group. The two groups were matched for age, gender, and educational level, and were found to be comparable (all *p*s > .24) in both the cognitive and affective dimensions of empathy (as measured by the Interpersonal Reactivity Index; (Davis, 1983), an important dispositional trait strongly affecting utilitarian judgments in moral dilemmas (Choe & Min, 2011; Crockett, Clark, Hauser, & Robbins, 2010; Gleichgerrcht & Young, 2013; Sarlo et al., 2014). However, we cannot exclude that other basic individual differences have determined or modulated the decision to consider or not legal implications, as well as the obtained between-group differences in affective ratings and electrophysiological activity. In particular, between-group differences in variables related to the general category of emotional distress, such as anxiety and depression, and/or variables more specifically affecting sensitivity to punishment, such as neuroticism and psychopathy, might have influenced the results, suggesting cautious interpretation of the causal role played by legal concerns.

Moreover, the information on the legal consequences of actions was not truly manipulated. However, instructing participants in the Legal group that both the proposed options were legally prosecutable would have spoiled the response patterns for at least two reasons: (a) few, if any, participants would have accepted that non-utilitarian options (e.g., deciding not to pull the lever in the Trolley dilemma, or not to push the man off the bridge in the Footbridge dilemma) were legally prosecutable at all; (b) even assuming that, we did not have any reference to quantify and differentiate the punishment assigned to each option: indeed, it would have been implausible that the utilitarian option (e.g., deciding to push the man off the bridge) received the same (unspecified) punishment as the non-utilitarian option (e.g., deciding not to push the man off the bridge). It is worth noting that the three Italian legal experts that we have consulted on this issue provided conflicting advice on the prosecutability of the actions proposed in the dilemmas. Specifically, according to two criminal defense lawyers, neither the utilitarian nor the non-utilitarian option

proposed in the dilemmas would be legally prosecutable, as it would be possible to appeal to the ‘‘case of necessity’’ (see Note 1) and to the lack of ‘‘legal obligation to act’’, respectively. In contrast, according to an assistant district attorney general at the Supreme Court, the utilitarian action would be legally prosecutable as homicide, albeit with possible mitigating circumstances. Importantly, all the three experts recognized that Trolley- and Footbridge-type dilemmas do not differ in terms of legal implications.

2.5. Conclusion

We believe that the present work contributed to the relevant literature in demonstrating the role played by legal implications in reducing the subjective emotional impact and the conflict associated with decision-making in the context of moral dilemmas. While modulating both emotional experience and neural activity preceding the behavioral choices, such effects seem to be largely independent of the intentionality of the proposed violations and to apply to both Footbridge- and Trolley-type dilemmas. Thus, it is unlikely that the difference in utilitarian responses between Footbridge-type and Trolley-type dilemmas that has been consistently reported in the literature is due to differences in hypothesized legal consequences.

This study also showed that Footbridge-type dilemmas, as compared to Trolley-type dilemmas, elicit a more unpleasant emotional state during the decision, as indicated both by self-reported emotional evaluations (though this effect was only marginally significant) and by a bigger amplitude of the P260 component. However, in this study, as well as in previous research (Sarlo et al., 2014), the self-reported emotional valence and arousal did not show an association with the percentage of utilitarian choices. At least two different reasons can be hypothesized to explain this result. The first would be that, although Footbridge-type dilemmas elicit stronger emotional reactions than Trolley-type dilemmas, the magnitude of this reaction has no direct influence on the probability of rejecting the utilitarian option. Indeed, the stronger emotional reactions experienced during the decision and the rejection of the utilitarian option could be two separate and independent

consequences of the evaluation of the intentionality of an action. The second possibility would be that the association between emotional reactions and the rejection of the utilitarian option exists, but it is not captured by the (conscious) emotional evaluations of valence and arousal experienced during the decision, because choices are not influenced by immediate emotions experienced during the decision process, but rather by anticipated emotions that participants expect to feel as a consequence of their choice. The second study presented in this thesis will investigate the former hypothesis, whereas the third study will focus on the latter.

3. STUDY II: THE INFLUENCE OF TRAIT PSYCHOPATHY ON JUDGMENT AND CHOICES IN MORAL DILEMMAS

3.1. Introduction

As seen in the general introduction, the dual process model of moral judgment (Greene et al., 2004, 2001) hypothesizes that in sacrificial moral dilemmas, which depict the choice between sacrificing one person and let several people die, emotional processing plays a crucial role in influencing decisions and judgments toward the rejection of the utilitarian option (i.e., sacrificing one to save more), competing with a deliberate cost-benefit analysis. For instance, it has been consistently reported in the literature that individuals endorse less the utilitarian option in dilemmas in which the sacrifice of one person is a means to save a greater number of people (Footbridge-type dilemmas) than in dilemmas in which the sacrifice is just a side effect (Trolley-type dilemmas) (Cushman et al., 2006; Lotto et al., 2014; Moore et al., 2008; Schaich Borg et al., 2006). According to Greene et al.'s model (2004, 2001), this would be a consequence of sacrifice as a means being emotionally more aversive than sacrifice as a side effect. However, only very few studies used self-report measures to test whether individuals indeed experience a stronger emotional reaction in Footbridge-type dilemmas, with mixed results (Lotto et al., 2014; Sarlo et al., 2012). Moreover, killing someone as a means to an end is perceived to be more intentional than killing someone as a side effect (Cushman & Young, 2011). Thus, it is unclear how emotional processing interacts with the means vs side effect distinction in affecting choices in moral dilemmas. In particular, it is unclear whether individuals would still reject the utilitarian option in Footbridge-type dilemmas, if killing as a means did not cause a strong emotional reaction.

One way to test this hypothesis would be to investigate emotional reactivity, and the respective decision choices, to Footbridge- and Trolley-type moral dilemmas in individuals with high trait psychopathy, since they are known to have blunted emotional reactions (Dawel, O'Kearney, McKone, & Palermo, 2012), but preserved attribution of intentionality (Blair et al.,

1996; Young, Koenigs, Kruepke, & Newman, 2012). In more detail, psychopathy is a clinical construct characterized by emotional hyporeactivity on the one hand and by a propensity to immoral behavior on the other (Cleckley, 1976; Hare, 2003). Importantly, it is a dimensional construct that can be found in various degrees in the general population (Edens, Marcus, Lilienfeld, & Poythress, 2006; Hare, 2003), and thus its influence over emotional processes and moral behavior can be seen as a quantitative continuum rather than referring to discrete taxonomies. Psychopathy consists of two independent factors: the first, known as "primary psychopathy", "emotional detachment", or "callous and unemotional trait", is characterized by callousness, and lack of guilt, regret and empathy; the second, known as "secondary psychopathy" or "antisocial factor", is characterized by impulsivity, irresponsibility, and a tendency towards antisocial behavior (Barry et al., 2000; Hare, 2003; Skeem, Johansson, Andershed, Kerr, & Louden, 2007). It is widely accepted that the core features of psychopathy are better represented by primary psychopathy, which is supposed to result from an intrinsic deficit in emotional processing (Karpman, 1941; Koenigs, Kruepke, & Newman, 2010; Verona, Patrick, Curtin, Bradley, & Lang, 2004). In this regard, as the study reported in the present chapter concerns the influence of emotion on morality, we will consider only primary psychopathy as pertinent, and in the rest of the thesis we will use the term "psychopathy" to refer to primary psychopathy.

Although there is a large consensus on the fact that individuals with psychopathy are characterized by emotional deficits, the specificity of such deficits is still under research. Some studies reported a specific deficit for fear-related stimuli (e.g., Marsh et al., 2011), whereas other studies reported a more widespread impairment extending to various negative and positive emotions (see Dawel et al., 2012), for a review). Nevertheless, there is consistent evidence that individuals with psychopathy show impaired recognition of distress cues and a diminished reactivity to stimuli such as facial or vocal expressions of fear, sadness and pain, and to pictures of wounded bodies (Blair, Jones, Clark, & Smith, 1997; Blair et al., 2013; Levenston et al., 2000). In particular, impaired recognition of facial expressions of fear, sadness, and surprise (Marsh & Blair, 2008),

selectively impaired recognition of fearful vocal affect (Blair, 2005), lack of startle blink potentiation during the viewing of high arousal negative pictures (Patrick, 1994), and reduced skin conductance reactivity to human emotional sounds (Verona et al., 2004) have been reported in the relevant literature.

Several studies have already investigated how psychopathy modulates responses to sacrificial moral dilemmas. The results show that, when confronted with Footbridge- and Trolley-type dilemmas, individuals with high psychopathy traits report they would perform the action of sacrificing one person to a greater extent than individuals with low psychopathy traits (Bartels & Pizarro, 2011; Glenn et al., 2010; Koenigs et al., 2012; Tassy, Deruelle, et al., 2013), a result that is consistent with the assumption that the aversive emotional response to harmful acts is weak or even absent in these participants. Moral judgments, instead, do not seem to be affected by psychopathy levels, as participants with high trait psychopathy do not judge sacrificing one person as more morally acceptable than participants with low trait psychopathy (Cima, Tonnaer, & Hauser, 2010; Glenn, Raine, Schug, Young, & Hauser, 2009; Tassy, Deruelle, et al., 2013; but see Patil, 2015). Such results could be explained by hypothesizing that, with no emotional response informing on the rightness of an action, individuals with psychopathy would produce altered choice of action, but normal moral judgment through preserved perspective-taking processes (Glenn, Raine, Schug, et al., 2009; Tassy, Deruelle, et al., 2013). Thus, these studies show a dissociation between moral *judgment* and *choice of action.* Indeed, judgments are made on a third-person perspective, thus relying on an allocentric reasoning stance (i.e., the situation is represented independently of one's current relation with it; (U. Frith & de Vignemont, 2005), whereas choices of action are made on a first-person perspective and are based on an egocentric point of view (Tassy, Oullier, Mancini, & Wicker, 2013). In addition, judgments and choices of action seem to rely on at least partially different neural substrates, as judgments, but not choices, were found to rely on processing in the right dorsolateral prefrontal cortex (Tassy et al., 2012). On the other hand, the fact that in psychopathic individuals choice of action was found to be altered in the face of unaffected moral

judgment suggests that choice of action is influenced by emotional processing to a greater extent than judgment.

It is important to note, however, that the above studies did not include any measure of the emotional state experienced by participants, and consequently did not provide evidence on the critical role of emotion in modulating this dissociation. Moreover, these studies were not aimed at testing whether the difference in choices between Footbridge- and Trolley-type dilemmas is due to differences in the engagement of emotional processing or to differences in attribution of intentionality alone. Indeed, these studies either use just one category of dilemmas (Bartels & Pizarro, 2011; Tassy, Deruelle, et al., 2013), did not directly compare Footbridge- and Trolley-type dilemmas (Seara-Cardoso et al., 2013, 2012), or did not use dilemmas that vary on the means/side effect distinction (Cima et al., 2010; Gao & Tang, 2013; Glenn et al., 2010; Glenn, Raine, & Schug, 2009; Koenigs et al., 2012).

Starting from these premises, the main aim of the present study was to investigate the role of emotional processes in moral judgment and decisions in sacrificial moral dilemmas. In particular, we aimed at testing whether the emotional state experienced during decision-making is associated with subsequent choices and judgments, and whether trait psychopathy is associated to a lower emotional reactivity to the dilemmas, beside producing more utilitarian choices of action, as suggested in the literature. Moreover, we aimed at testing how the emotional hyporeactivity to harm that characterizes individual with high trait psychopathy affects choices and judgments in dilemmas that differ on the means vs side effect distinction. Participants' affective state during decision-making was assessed along the independent dimensions of valence (pleasantness/unpleasantness) and arousal (activation/calm), as they represent the core affective components of emotion (e.g., (Barrett & Fossum, 2001; Lang et al., 2008). This allowed to clarify whether emotion plays a crucial role in moral decision-making specifically when harm is a means to an end. Furthermore, by measuring both moral judgment and choice of action, we were interested in disentangling whether emotion differentially affects these two distinct processes underlying moral behavior.

3.2. Methods

3.2.1. Participants

The Levenson Self-Report Psychopathy Scale (LSRP, Levenson, Kiehl, & Fitzpatrick, 1995), a self-assessment questionnaire measuring primary and secondary psychopathy, was administered to 281 university students. From this sample, 54 subjects were selected on the basis of their scores on the *primary psychopathy* scale (LSRP-1), as the emotional hyporesponsiveness characterizing individuals with psychopathy is linked entirely to this factor (e.g., Vanman, Mejia, Dawson, Schell, & Raine, 2003). Twenty-six participants who scored equal to or higher than 32 (i.e., the 75th percentile of the distribution) were included in the *High Psychopathy* (HP) group. Twenty-eight participants who scored equal to or lower than 24 (i.e., the 25th percentile) were included in the *Low Psychopathy* (LP) group. Three participants of the LP group were excluded because of technical problems during the experiment or because of non-compliance with the instructions. The final group was thus composed of 25 participants. The two groups were comparable for age and gender (HP: 15 F, 11 M, mean age = 22.08 years, *SD* = 1.13 years; LP: 15 F, 10 M, mean age = 21.32 years, *SD* = 1.55 years). All participants read and signed an informed consent. The study was conducted in compliance with the declaration of Helsinki on research on human subjects and was approved by the local ethical committee.

3.2.2. Stimuli

The stimuli consisted of 28 moral dilemmas taken from a standardized set (Lotto et al., 2014), that described hypothetical scenarios in which the agent must choose whether or not to kill one individual to save more people. In half of the dilemmas, killing one individual was a foreseen but unintended consequence of saving others (Trolley-type dilemmas), whereas in the other half, killing one individual was an intended means to save others (Footbridge-type dilemmas). The

stimuli were presented on a 19 inch. monitor at a viewing distance of 100 cm. Stimulus presentation was accomplished with E-prime software (Psychology Software Tools Inc., 2012).

3.2.3. Procedure

The stimuli were divided in 4 blocks and presented in a pseudo-randomized order. In each trial, two text slides describing the scenario and the action that could be performed were presented. Participants read the slides at their own pace and advanced by pressing the spacebar. After reading the two slides, they rated to what extent they would perform the action (*choice of action*) on a scale ranging from 0 (absolutely not) to 7 (absolutely yes). Immediately after, they rated the emotional state they had experienced during the decision, using the valence and arousal scales of the Self-Assessment Manikin (Lang et al., 2008). The valence scale ranged from extreme unpleasantness (1) to extreme pleasantness (9); the arousal scale ranged from extreme calm (1) to extreme excitement (9). Finally, participants rated to what extent the proposed action was morally acceptable (*moral judgment*) on a scale from 0 (not at all) to 7 (completely). Between trials, a fixation cross was presented on screen for 1000 ms. The task lasted about thirty minutes.

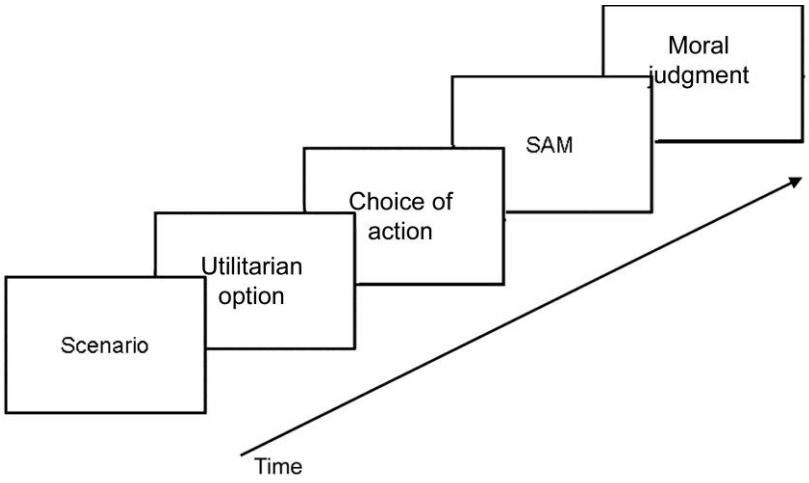


Figure 3.1. Sequence of events in the experiment. Participants had to rate to what extent they would perform the action (choice of action) on a scale ranging from 0 (absolutely not) to 7 (absolutely yes), and to what extent the proposed action was morally acceptable (moral judgment) on a scale from 0 (not at all) to 7 (completely). Participants could read at their own pace and all slides remained on screen until keypress. SAM = self-assessment manikin.

3.2.4. Statistical analysis

The mean scores of valence, arousal, choice of action, and judgment were calculated for each participant and each type of condition. These scores were entered separately in two 2x2 Analyses of Variance (ANOVAs) with group (HP vs. LP) as between-participants factor and dilemma type (Trolley-type vs. Footbridge-type dilemmas) as within-participants factor. Significant interactions were further analyzed with Tukey post-hoc tests.

To test the association between emotional state, action choice and moral judgment, for each group and dilemma type we performed Pearson's correlations between arousal and valence scores and choice of action and judgment scores separately. All analysis were performed using IBM'SPSS (IBM Corp., 2010).

3.3. Results

3.3.1. Choice of action

The group main effect was significant ($F(1,49) = 15.33$, $p = .02$, $\eta^2 = .24$,), with participants in the HP group more inclined to sacrifice one person to save a larger number of people in both types of dilemmas as compared to participants in the LP group (Figure 3.2). The dilemma type main effect was also significant ($F(1,49) = 127.42$, $p < .001$, $\eta^2 = .72$), as participants were more inclined to sacrifice one person in Trolley-type as compared to Footbridge-type dilemmas (*M*s = 3.53 and 1.29, respectively). The Group x Dilemma type interaction was not significant.

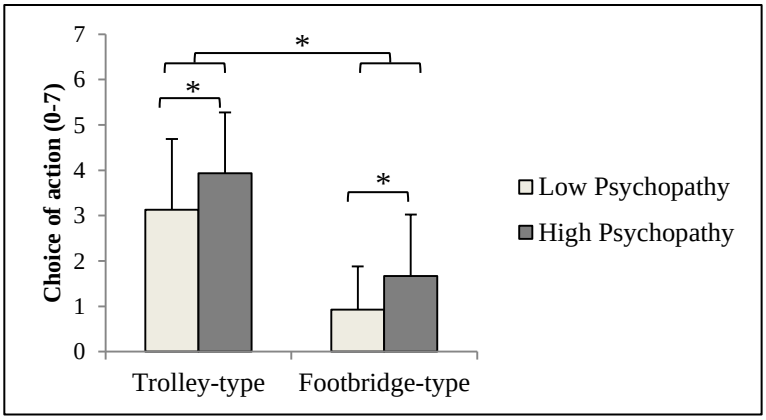


Figure 3.2. Mean ratings for choice of action in low and high trait psychopathy groups. The scale ranged from 0 (absolutely no) to 7 (absolutely yes). Error bars indicate the standard errors.

3.3.2. Moral judgment

The significant dilemma type main effect ($F(1,49) = 58.25$, $p < .001$, $\eta^2 = .54$) showed that participants judged sacrificing one person in Footbridge-type dilemmas as less morally acceptable than in Trolley-type dilemmas (Figure 3.3). No other effect was significant.

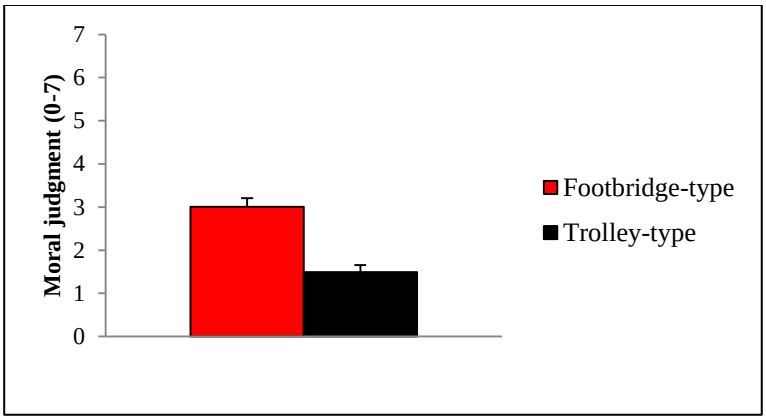


Figure 3.3. Mean ratings for moral judgment in low and high trait psychopathy groups. The scale ranged from 0 (not at all morally acceptable) to 7 (completely morally acceptable). Error bars indicate the standard errors.

3.3.3. Emotional reactivity

3.3.3.1. Valence

The significant group main effect ($F(1,49) = 21.39$, $p = .01$, $\eta^2 = .30$), showed that participants in the HP group reported lower unpleasantness (i.e., higher valence scores) during decision as compared to participants in the LP group (Figure 3.4). No other effects were significant.

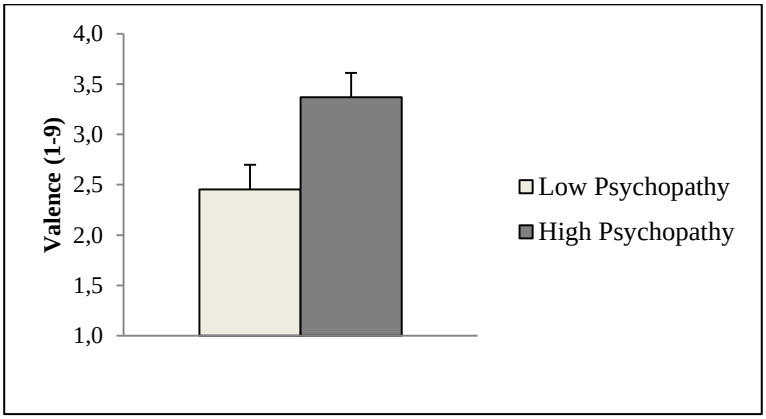


Figure 3.4. Mean ratings for valence experienced during decision-making in low and high trait psychopathy groups. The scale ranged from 1 (estreme unpleasantness) to 9 (extreme pleasantness). Error bars indicate the standard errors.

3.3.3.2. Arousal

The significant dilemma type main effect ($F(1,49) = 1.33$, $p = .02$, $\eta^2 = .03$), indicated that decisions in Trolley-type dilemmas were more arousing than in Footbridge-type dilemmas (Figure 3.5). No other effects were significant.

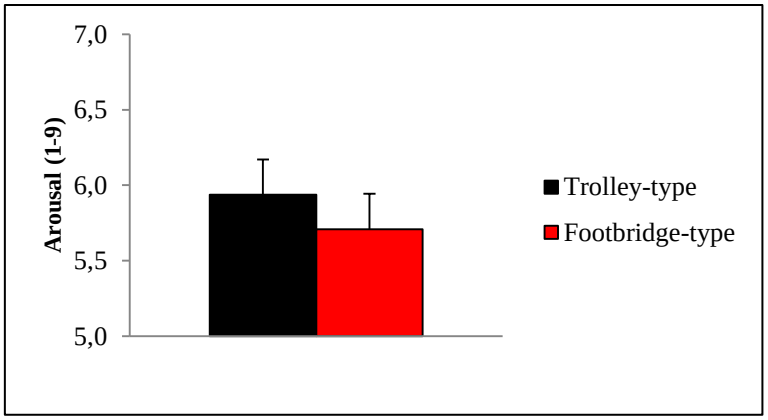


Figure 3.5. Mean ratings for arousal experienced during decision-making in low and high trait psychopathy groups. The scale ranged from 1 (extreme calm) to 9 (extreme excitement). Error bars indicate the standard errors.

3.3.4. Correlations

In the HP group, the Valence experienced during Trolley-type dilemmas was positively correlated with Judgment scores(r=.44, p=.03). None of the other correlations was significant.

3.4. Discussion

The aim of this research was to investigate whether and how emotional hyporeactivity, as typical of individuals high in trait psychopathy, affects moral judgments and choices of action in sacrificial moral dilemmas. We hypothesized that higher levels of psychopathy would be associated with lower emotional reactivity to harm and with a higher propensity to endorse acts causing harm to others. At variance with previous studies, by employing a reliable measure of the emotional state experienced by participants we were able to provide empirical evidence to test this hypothesis. Moreover, we investigated whether a stronger emotional reactivity to intentional harm as compared to unintentional harm is necessary to reject the utilitarian option in Footbridge-type dilemmas more than in Trolley-type dilemmas.

We found that HP participants, compared to LP participants, were more likely to act sacrificing one person (i.e., choice of action), regardless of whether this sacrifice was a means (i.e., Footbridge-type dilemmas) or a side effect (i.e., Trolley-type dilemmas) of saving more people.

This is in line with what reported on individuals with psychopathy in previous studies using sacrificial moral dilemmas (Bartels & Pizarro, 2011; Glenn et al., 2010; Koenigs et al., 2012; Tassy et al., 2013). This finding has been interpreted in the literature as reflecting a reduced emotional reaction to the idea of harming someone, which would make the rational choice of maximizing the outcomes to emerge (Tassy, Deruelle, et al., 2013). Consistent with this view, we found that HP participants reported less unpleasantness during decision-making as compared to LP participants. In accordance with the dual process model of moral judgment (Greene et al., 2004, 2001), these results indicate that a lower emotional reactivity during decision-making, as measured by subjective evaluations, is associated with a higher disposition to endorse utilitarian choices in individuals with high psychopathy.

Interestingly, no difference between groups emerged in moral judgment, with both the HP and the LP groups judging Footbridge-type dilemmas as less morally acceptable than Trolley-type dilemmas, as consistently found in the relevant literature (Cushman et al., 2006; Lotto et al., 2014; Moore et al., 2008; Schaich Borg et al., 2006). The finding of a lack of behavioral differences in moral judgment, consistent with what reported in recent studies (Cima et al., 2010; Glenn, Raine, Schug, et al., 2009; Tassy et al., 2013), highlights a dissociation between moral judgment and choice of action, suggesting the existence of different underlying processes. For instance, recent studies using sacrificial moral dilemmas reported a discrepancy between judgment and choice of action, showing that people do not always choose the action that they deem morally appropriate (Kurzban, DeScioli, & Fein, 2012; Tassy, Oullier, et al., 2013). Moreover, choices of action, as compared to judgments, are more strongly influenced by personally relevant variables, such as the affective proximity with the victim (Kurzban et al., 2012; Tassy, Oullier, et al., 2013). This dissociation can be explained considering that moral judgments are made from an allocentric perspective and might rely more strongly on normative prescriptions and beliefs (Nichols & Mallon, 2006). In contrast, choices of action elicit more self-referent processing (Sood & Forehand, 2005), have personally relevant consequences (Sood & Forehand, 2005), and are made by taking personal

experiences and motivations into greater account (Zeelenberg, 2008). Thus, it is possible that choices of action are strongly affected by the emotions elicited when people consider the alternatives and the expected outcomes (Loewenstein & Lerner, 2003; Tassy, Oullier, et al., 2013; Zeelenberg, 2008).

Indeed, judgment can be driven by moral knowledge learned through other processes, such as explicit rule learning or observational learning (Darley & Shultz, 1990). On this basis, the HP participants provided judgments of moral acceptability similar to those of the LP participants, being able to differentiate between situations in which harming others is intentionally caused and situations in which harm is a side effect. This is consistent with the findings reported by Young and colleagues (Young et al., 2012), in which psychopathic individuals, presented with a series of vignettes describing intentional, attempted, and accidental harms, based their moral judgments on the actors' intents and judged unintentional harm as more acceptable than both intentional and attempted harms. This indicates that coding an action as intentional and judging an intentional damage as less acceptable than an unintentional one do not necessarily require emotional processing. Indeed, the ability of understanding someone else's intentions is thought to be based on the Theory of Mind abilities (Frith & Frith, 2005), which are not impaired in individuals with psychopathy (e.g., Blair et al., 1996).

As regards the self-evaluations of the emotional state experienced during the dilemmas, it is important to note that, against the expectations, no significant difference between the valence reported for Trolley-type and Footbridge-type dilemmas emerged, as well as no significant correlation between valence and choice of action in either of the two groups. The only significant correlation reported between emotional ratings and responses to the dilemmas was the positive correlation between valence and moral judgment ratings that the HP group showed for Trolley-type dilemmas. This result alone doesn't provide a strong evidence supporting the hypothesis of an association between emotion and moral judgment. Moreover, a result that may seem surprising is the fact that Trolley-type dilemmas were rated as more arousing than Footbridge-type dilemmas.

This may seem contradictory, as Footbridge-type dilemmas are supposed to be more emotionally salient than Trolley-type dilemmas (e.g., Sarlo et al., 2012), and thus could be expected to elicit heightened arousal. However, arousal *per se* could result both from emotional activation and from cognitive engagement (Critchley, Corfield, Chandler, Mathias, & Dolan, 2000). This last explanation is probably the most suitable in this case: we found no difference in arousal between groups, but we found that HP participants reported reduced unpleasantness while deciding in moral dilemmas as compared to LP participants. Thus, there is evidence that HP participants had a reduced emotional reaction to the dilemmas. The equally arousing state experienced by the two group might be more related to cognitive effort, which is supposedly higher in Trolley-type dilemmas (Greene et al., 2001).

3.5. Conclusion

In conclusion, the present study supports and extends previous research in demonstrating that high psychopathy trait reduces the emotional reactivity to acts involving physical harm to others and increases utilitarian responses in sacrificial dilemmas. Furthermore, the observed dissociation between choice of action and moral judgment, which has been previously demonstrated, suggests that choice of action is more closely related to emotional experience, whereas judgment mainly relies on cognitive perspective-taking processes. However, our studies do not provide further support to the hypothesis that emotional processing drives moral decisions, at least as far as concerns the emotions of which participants are aware, because the association between emotional self-reports and choices never reached significance. Moreover, the unpleasantness experienced by participants during the decision was comparable for the two dilemma types.

Importantly, our results indicate that a stronger emotional reaction to intentional harm may not be necessary to reject the utilitarian resolution in Footbridge-type dilemmas, as opposed to what hypothesized by the dual process model (Greene et al., 2004, 2001; Greene, 2008). In fact, the

influence of high trait psychopathy on choices equally affected both dilemma types and the difference between the two dilemma types in choices of action and judgments was constant in the two groups: both groups judged the utilitarian option in Footbridge-type dilemmas as less morally acceptable than in Trolley-type dilemmas, and both groups reported to be more inclined to perform the action described in utilitarian option in Trolley- than in Footbridge-type dilemmas. Taken together, these results seem to indicate that the difference in choices and judgments between Trolley- and Footbridge-type dilemmas could not only be due to differences in emotional processing, but also to other elements, like the appraisal of a difference in intentionality.

As an alternative, it can be hypothesized that the influence of emotional processing on decisions is not captured by the immediate emotions that participants subjectively experience during the decision. For instance, several models of decision-making (Bell, 1982, 1985; Loomes & Sugden, 1982, 1986; Mellers, Schwartz, & Ritov, 1999) propose that decisions are driven by the need to minimize the expected emotions associated with the outcomes of the decision, more than by the immediate emotions experienced during the decision. This hypothesis will be addressed in the third study.

4. STUDY III: THE ROLE OF ANTICIPATED EMOTIONS IN THE RESOLUTION OF MORAL DILEMMAS

4.1. Introduction

As seen in the previous chapters, according to Greene and colleagues' dual process model (2004, 2001), moral judgments and decisions are driven by two systems in competition: a deliberate and rational one and an automatic and emotional one. Starting from this model, several studies tested the hypothesis of emotional processing having a major role in moral decisions and judgments, yielding generally consistent, but not conclusive, results. As seen in the general introduction, the resolution of Footbridge-type dilemmas elicits greater activity in brain areas associated with emotional processing, like the ventromedial prefrontal cortex (vmPFC), the superior temporal sulcus (STS) and the amygdala, as compared to Trolley-type dilemmas (Greene et al., 2004, 2001). Furthermore, individuals with emotional hyporeactivity (e.g., participants with high psychopathy traits, or vmPFC impairments) show a higher endorsement of utilitarian options as compared to control participants (e.g. Koenigs et al., 2007, 2012; Moretto et al., 2010), a result that was also replicated in the study presented in Chapter 3 of this thesis.

Despite this evidence, however, up to now no study reported a significant association between self-report ratings of the emotional state experienced during decision-making and moral judgments or choices. Furthermore, the studies reported in Chapter 2 and Chapter 3 of this thesis did not find any significant correlation between the emotional state experienced during the decision and the percentage of utilitarian choices. Thus, it is still unclear whether and how decision-making in moral dilemmas is affected by the emotions that individuals experience relative to their resolution.

One important issue that may help to shed light on these incongruences is the fact that the literature on moral dilemmas has mainly focused on immediate emotions that participants experienced at the moment of the decision, and has not examined anticipated post-decisional emotional consequences. This is a relevant point, because according to several models of decision-

making like the regret theory (Bell, 1982; Loomes & Sugden, 1982), the disappointment theory (Bell, 1985; Loomes & Sugden, 1986) and the decision affect theory (Mellers et al., 1999), the anticipation of post-decisional emotions has a crucial impact on decisions. According to these models, individuals try to predict how they would feel after having chosen each of the different alternatives the problem they are faced with entails, and then select the option that minimizes the anticipated negative emotions. In particular, according to these models, individuals are especially motivated to avoid post-decisional feelings of regret arising when the outcome of the chosen option is worse than what would have resulted from the alternative options. Additionally, as concerns the moral domain, individuals are also motivated to avoid anticipated feelings of guilt. Guilt is a negative emotion elicited by causing harm or distress to others, and entails self-condemning feelings (Haidt, 2003; Tangney et al., 1996). Several studies indicate that the need to avoid guilt strongly influences decision-making when the individual's decisions have relevant consequences for others. In particular, guilt aversion motivates individuals to act cooperatively (Chang, Smith, Dufwenberg, & Sanfey, 2011; de Hooge et al., 2007; Ketelaar & Tung Au, 2003) and to avoid deception (Charness & Dufwenberg, 2006).

It is plausible to hypothesize that also in moral dilemmas, decisions would be driven by the attempt to minimize post-decisional negative emotions. Coherently with this hypothesis, a recent study (Sarlo et al., 2014) reported that the percentage of utilitarian choices made in moral dilemmas was inversely related to the individuals' disposition to experience personal distress when faced with the suffering of others. This suggests that anticipating the personal distress that would result after killing someone might motivate individuals to reject the utilitarian option.

There are several distinct emotions that can be hypothesized to arise from choosing either the utilitarian or the non-utilitarian option in moral dilemmas. For instance, sacrificing one person could elicit more regret and guilt than letting several people die, because regret and guilt are enhanced by feelings of personal responsibility (e.g.,(Wagner et al., 2012), which in turn are enhanced by action as compared to inaction (Zeelenberg, van der Pligt, & de Vries, 2000). Another

emotion that can be hypothesized to play a role in moral dilemmas is shame, a moral emotion that is elicited by the violation of a moral rule or a social standard and that, as opposed to guilt, focuses on self-image rather than on the outcomes of an action (e.g., Haidt, 2003; Tangney et al., 1996). Since both of the options depicted in the dilemmas entail the violation of a moral rule or standard (killing in the case of the utilitarian option, not helping people in danger in the case of the non-utilitarian option), we can hypothesize both options to elicit feelings of shame. Finally, two additional emotions that can be hypothesized to play a role are anger and disgust, since a study by Choe and Min (2011) found that trait dispositions to experience those emotions were associated with the percentage of utilitarian judgments. In more detail, high trait anger was associated with a higher percentage of utilitarian judgments, whereas high trait disgust was associated with a lower percentage of utilitarian judgments. In line with these findings, a study by Ugazio, Lamm and Singer (2012) showed that inducing anger before a moral dilemma task increased the endorsement of the utilitarian option, whereas inducing disgust reduced it. Both anger and disgust are basic emotions that are also elicited by moral violations (Haidt, 2003). In particular, moral anger is caused by perceived violations of a person's freedom, property, or rights (Rozin, Lowery, Imada, & Haidt, 1999); moral disgust is elicited by violation of norms concerning purity, sanctity, and the natural order of things (Rozin et al., 1999). Disgust can also be directed to the self, being in this case similar to shame (Simpson, Hillman, Crawford, & Overton, 2010).

A secondary goal of this study was to investigate the electrophysiological correlates of the last phase of the decision-making process, in which an option is selected and the corresponding action is implemented. In particular, we were interested in the readiness potential (RP), a slow negative wave that, as seen in Chapter 2, is observed before the execution of a voluntary movement. The RP recorded over the central electrode Cz reflects an increase in the cortical excitability of brain areas involved in the preparation of the movement, like the Supplementary Motor Area (SMA) and the Pre-motor Cortex (Shibasaki & Hallett, 2006). Apart from being involved in the preparation and selection of actions (Rushworth, Walton, Kennerley, & Bannerman, 2004), the

SMA plays a crucial role in value-based decision making, being involved in reward anticipation (Lee, 2004) and in the encoding of the reward value associated with an action (Wunderlich et al., 2009).

In line with these functions of the SMA, some recent findings indicate that the RP is a relevant correlate of decision-making. For instance, a recent study found that the readiness potential tracks the emergence of value-based decisions and reflects the readiness to decide (Gluth et al., 2013). In that study, the neural correlates of value-based decisions were investigated using a sequential decision making paradigm in which participants could either buy or reject stock offers based on probabilistic information about the stocks' value that were delivered through sequentially presented ratings. Participants had to pay a price for each of these ratings, and thus the optimal strategy was to respond as soon as the minimum amount of information necessary for the choice was collected. The cost of these ratings was manipulated to create a high-cost condition and a low-cost condition. As a result, a greater amplitude of the RP was observed for the high-cost condition as compared to the low-cost condition. Moreover, the amplitude of the RP measured at Cz increased with time, so that early rating presentations were associated with a weaker signal than late rating presentations. Thus, the readiness potential reflected the urge to decide, that gradually increased together with the motivation to respond.

The RP seems to have an important role also in moral decisions, since it might reflect moral conflict. As we saw in the introduction and in Chapter 2, lower amplitudes of the RP were reported for Footbridge-type as compared to Trolley-type dilemmas, a result that was interpreted as reflecting a lower preparation to respond, probably due to a greater conflict between alternative options (Sarlo et al., 2012). In coherence with this finding, an fMRI study reported greater SMA activation for Trolley-type as compared to Footbridge-type dilemmas (Schaich Borg et al., 2006). Moreover, as seen in Chapter 2, taking into account legal consequences during the resolution of moral dilemmas was associated with a greater RP amplitude as compared to not taking legal consequences into account. Since not taking legal consequences into account makes the resolution

of moral dilemmas more unpleasant and presumably more difficult, because individuals cannot use legal consequences as a reference point, the lower amplitude of the RP was interpreted as a sign of conflict in this case as well as in the study by Sarlo and colleagues (2012). Finally, another EEG study found a smaller RP amplitude for deception than for truth telling (Panasiti et al., 2014), which can also be related to moral conflict, this time between a personal advantage and a morally correct alternative.

Taken together, these results suggest a relationship between the amplitude of the readiness potential and the intensity of conflict in moral situations, but this interpretation requires further testing, and what exactly the RP reflects in morally conflictual situations has still to be investigated. Thus, in the present experiment we aimed at testing if the RP reflects a conflict between anticipated emotional consequences of the two alternative options of the dilemmas, hypothesizing that a greater conflict between alternatives would reflect in a lower readiness to decide. To this aim, we measured emotional experiences related to both options for each dilemma, and we used this data to calculate an index of emotional conflict between the two options, testing its association with the RP amplitude.

To summarize, the main aim of this study was to test the influence of anticipated emotional consequences on decisions in moral dilemmas. We hypothesized participants to choose the option associated with a lower emotional cost. Furthermore, we hypothesized that in Footbridge-type dilemmas, but not in Trolley-type dilemmas, the utilitarian option would be associated with worse anticipated emotional consequences as compared to the non-utilitarian option. This would help to explain why in Footbridge-type dilemmas people reject the rational option that maximizes the number of saved lives, whereas in Trolley-type dilemmas people endorse it. We focused on six different emotions: first of all regret, that is an important decision-related emotion elicited by comparing the outcome of choices. In Italian, regret can be translated in three main words: *rimorso*, which is more related to action; *rimpianto*, which is more related to inaction and *rammarico*, which is generally related to bad outcomes (Giorgetta, Zeelenberg, Ferlazzo, & D'Olimpio, 2012). We

chose to focus on the first two emotional labels, because they are differentially influenced by agency and responsibility (Giorgetta et al., 2012), and thus can be hypothesized to be differentially elicited by the utilitarian and non-utilitarian options in moral dilemmas. Then, we chose to measure guilt and shame, that are perhaps the two most prototypical moral emotions (Haidt, 2003), and disgust and anger, that are the only two basic emotions that can be elicited by moral violations (Haidt, 2003). For each dilemma, we collected self-report measures of the emotional state experienced by participants relative to both the utilitarian and the non-utilitarian option, and we tested whether participants chose the option associated with the lower emotional cost.

Finally, the second aim of this study was to further investigate the role of the RP in this context by testing if it reflects a conflict between anticipated emotional consequences associated with the two choice options.

4.2. Methods

4.2.1. Participants

Sixty healthy participants (30 women) aged 19–26 years (*M*= 22.3, *SD* = 1.71) took part in the present study. All participants were right-handed and had no history of psychiatric or neurological disorders, and they were paid €13 for their participation. Of them, 4 participants were excluded due to technical problems, 3 for non-compliance with the instruction, and 2 because they had already a previous knowledge of moral dilemmas. The final sample for behavioral data was thus composed of 51 participants (29 F).

As concerns EEG data, 8 additional participants were excluded because of excessive EEG artifacts, and the final sample was composed of 43 participants (23 F). To compute the Movement Related Potentials (MRPs), a minimum of 30 trials for condition is needed. Since the task would have been excessively long by administering 60 dilemmas to each participant (see procedure), we divided participants into two groups. One group (the *Footbridge* group) was administered 30 Footbridge-type dilemmas and 10 Trolley-type dilemmas, the other group (the *Trolley* group) 30

Trolley-type dilemmas and 10 Footbridge-type dilemmas. The two groups were comparable for age and gender (Trolley group: 11 F, 10 M, mean age = 22.9 years, *SD* = 1.41 years; Footbridge group: 12 F, 10 M, mean age = 22.27 years, *SD* = 1.69 years). For the MRP analysis, only Footbridge-type dilemmas were considered for the Footbridge group and vice versa for the Trolley group. For the behavioral data, on the other hand, all dilemmas were included in the analysis for every participant.

4.2.2. Stimuli

A set of 60 standardized dilemmas (Lotto et al., 2014) was used in this study. The set includes 30 Footbridge-type dilemmas, which described killing one individual as an intended means to save others, and 30 Trolley-type dilemmas, which described killing one individual as a foreseen but unintended consequence of saving others. All dilemmas were presented as written text on three consecutive slides: the *scenario* slide described the context, in which a threat endangers several people's lives; Option A described the non-utilitarian choice, in which the agent lets these people die; Option B described the utilitarian choice, in which the agent kills one person to save these people. Additionally, two *counterfactual* slides were created for each dilemma. These slides described in neutral terms the consequences of the alternative option in terms of number of deaths and lives. For instance, the counterfactual slide for Option A stated that, if the participant had chosen to perform the action described in Option B, one person would be dead and *n* people would be alive. The counterfactual slide for Option B stated that, if the participant had not chosen to perform the action described in the option, *n* people would be dead and one person would be alive.

Twelve additional moral dilemmas, which involved no deaths and described other moral issues (e.g., stealing, lying, and being dishonest), were used as filler stimuli.

Stimulus presentation was accomplished with E-prime software (Psychology Software Tools, Pittsburgh, PA).

4.2.3. Procedure

Upon arrival at the laboratory, participants read and signed an informed consent form. After an elastic cap was applied for EEG recording, instructions for the task were given. As already stated above, we chose not to administer the whole dilemma set to each participant, because the task would have lasted more than two hours and would have been excessively repetitive and tiring for the participants, thus compromising the reliability of their performance. For this reason, participants in the Footbridge group were presented with 30 Footbridge-type dilemmas, 10 Trolley-type dilemmas, and 4 fillers to avoid automaticity in the responses; participants in the Trolley group were presented with 30 Trolley-type dilemmas, 10 Footbridge-type dilemmas, and 4 fillers.

Dilemmas were presented in a pseudo-randomized order, so that each of the 2 experimental blocks was comprised of 15 dilemmas of the main category for the group, 5 dilemmas of the other category, and 2 fillers. In each trial, participants read the three text slides describing the scenario and the two options at their own pace and advanced by pressing the spacebar. Then, a fixation cross appeared on screen, and participants were instructed to decide between the two options by pressing one of two computer keys marked ‘‘A’’ and ‘‘B’’ with the index or the middle finger of the right hand. The fixation cross remained on screen until participants responded, for a maximum time of 10 seconds, plus one additional second after the response to prevent the MRPs to be contaminated by offset-related potentials. Then, participants rated how they felt after the decision on six 0-6 Likert scales indicating the intensity of 6 emotions: anger, disgust, guilt, *rimorso* (action regret), *rimpianto* (inaction regret), and shame. Subsequently, participants read the counterfactual slide relative to the option chosen and rated how they would have felt if they had taken the alternative option on the same 6 emotional scales. The emotions were presented in a random order. The stimuli were displayed on a 19 inch. monitor at a viewing distance of 100 cm, and the experimental task started after three practice trials. The task lasted about one hour, plus half-an-hour preparation time.

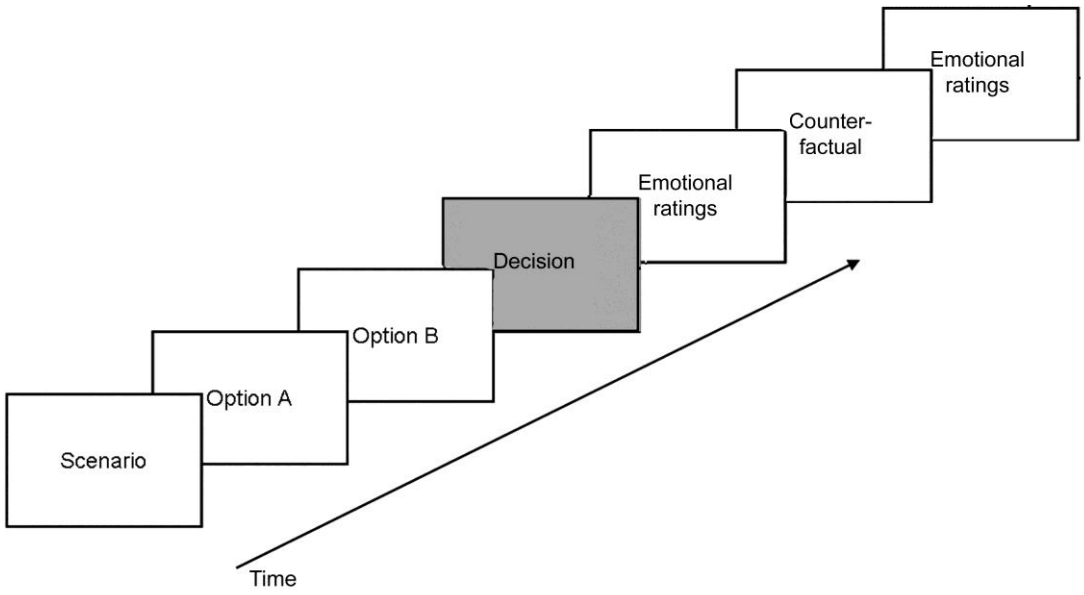


Figure 4.1. Sequence of events in the experiment. Participants had to decide between Options A and B by pressing the corresponding key during the presentation of the decision slide (in gray). During the first emotional ratings they had to rate how they felt after having chosen. Then, they were presented with a counterfactual slide describing what would have happened if they had chosen the other option. In the second emotional ratings, they had to rate how they would have felt if they had chosen the other option. MRPs were recorded time-locked to the behavioral response, during the decision slide. Text is not drawn to scale.

4.2.4. Data Collection and Analysis

Behavioral data were analyzed with mixed effect models, including all trials for each participants except those pertaining to the filler dilemmas, those in which participants didn't respond in time (N = 33, maximum number per participant = 6), and one trial with a response time < 120 ms.

First of all, to investigate if the emotions associated with the utilitarian and non-utilitarian options differed, and if this difference was modulated by dilemma type, we built a separate mixed effect linear regression model for each emotion, with emotional intensity as dependent variable, option to which the emotion was associated (utilitarian, non-utilitarian), dilemma type (Trolley-type, Footbridge-type) and Option × Dilemma Type as fixed effects, and participant and dilemma as random effects. To investigate if the dilemma type influenced the probability of choosing the utilitarian option, we built a mixed effect logistic regression model with choice (0 = non-utilitarian, 1 = utilitarian) as dependent variable, dilemma type as fixed effect, and participant and dilemma as random effects. Finally, to investigate if participants chose the option that was associated with the

lowest emotional intensities and if this effect was modulated by the dilemma type, we calculated for each trial and each emotion a differential intensity index by subtracting the intensities associated with the non-utilitarian option from those associated with the utilitarian option. We also calculated a mean emotional differential intensity index, to have an overall measure of emotional difference between the two options. Then, we built a separate mixed effect logistic regression model for each emotion with choice as dependent variable, differential emotional intensity, dilemma type, and Differential Emotional Intensity × Dilemma Type as fixed effects, and participant and dilemma as random effects. Importantly, for the behavioral data the dilemma type was a within participant factor, since we analyzed the data collected for both dilemma types for each participants

For each analysis, we started with the model including only the random effects and then introduced the fixed effects one by one, in the order described above. To compare models, we used the log-likelihood ratio test. To test the significance of parameters of the fixed effects, Wald z-tests were used for logistic models and t-test with the Satterthwaite approximations for degrees of freedom for linear models. As for electrophysiological data, the EEG was recorded from a tin electrode applied on the right mastoid and from 9 tin electrodes (Fz, Cz, Pz, F3, F4, C3, C4, P3, P4) embedded in an elastic cap (Electro-Cap International, Inc.; Eaton, OH). All impedances were kept below 10 kΩ, and the left mastoid was used as reference. After the recordings, all sites were re-referenced to the average of the left and right mastoids. Vertical and horizontal electro-oculogram were recorded from additional electrodes placed above and below the left eye and at the external canthi of both eyes, with the left mastoid as online reference and offline bipolar re-referencing. The signal was amplified with a BrainVision V-Amp amplifier (Brain Products GmbH, Gilching, Germany), bandpass filtered (DC- 70 Hz) and digitized at 500 Hz (24 bit A/D converter, accuracy 0.04 uV per least significant bit). Blink artifacts and eye movements were corrected with a regression-based algorithm (Gratton et al., 1983). The EEG was epoched into 1500-ms segments starting from 1000 ms before the keypress and ending 500 ms after. To correct for slow DC shifts, each epoch was linear detrended. Then, each epoch was re-filtered with a 30 Hz low pass filter (12

dB/oct) and baseline-corrected against the mean-voltage recorded during a 200 ms period preceding keypress (from -1000 ms to -800 ms) for MRPs. Only epochs pertaining Trolley-type dilemmas were retained for the Trolley group, and vice versa for the Footbridge group. The epochs were then visually screened for artifact and each epoch containing a higher voltage than ±70 μV in any channel was rejected from further analysis. The remaining epochs were averaged separately for each participant (mean retained epochs for the Trolley group: 23.14, SD 7.58; mean retained epochs for the Footbridge group: 22.98, SD 7.68). The amplitude of the readiness potential was measured in two time intervals (Shibasaki & Hallett, 2006): (1) mean negativity between 800 and 500 ms before keypress (early readiness potential); and (2) mean negativity between 500 and 50 ms before keypress (late readiness potential). Statistical analyses were then restricted to Cz since the RP measured at this electrode reflects the activation of the SMA (Shibasaki and Hallett, 2006), and since in the study by Gluth and colleagues (2013) the potential recorded at Cz tracked the emergence of value-based decisions.

First of all, t-tests with Welch-corrected degrees of freedom were performed separately for each time window to compare the amplitude of the readiness potential between dilemma types. Importantly, for all the analyses performed on RP amplitude, the dilemma type was a between-participants factor, since only Footbridge-type dilemmas were included in the analyses for the Footbridge group and only Trolley-type dilemmas were included in the analyses for the Trolley group. Then, to test if the amplitude of the readiness potential reflected emotional conflict and if this was modulated by dilemma type (i.e., by the group), an index of emotional conflict was calculated for each emotion, and for the mean of all emotions, as the absolute value of the difference between median emotional intensities associated to the two options. This was done in order to have an aggregate data to relate to the average RP amplitude. The median was chosen because it better reflected the center of the distributions of the emotional ratings, which were asymmetrical. Smaller values of these indexes of emotional conflict indicated similar intensities for the two options, and thus a higher emotional conflict; larger values indicated a stronger emotional

intensity for one of the two options, and thus a lower emotional conflict. Then, linear regressions were calculated separately for each emotion and time-window, with readiness potential amplitude as dependent variable and emotional conflict, dilemma type and the interaction between these two terms as predictors.

All statistical analysis were performed in R (R Core Team, 2015), using the libraries stats (R Core Team, 2015), lme4 (Bates, Mächler, Bolker, & Walker, 2014), lmerTest (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2015) and effects (Fox, 2003).

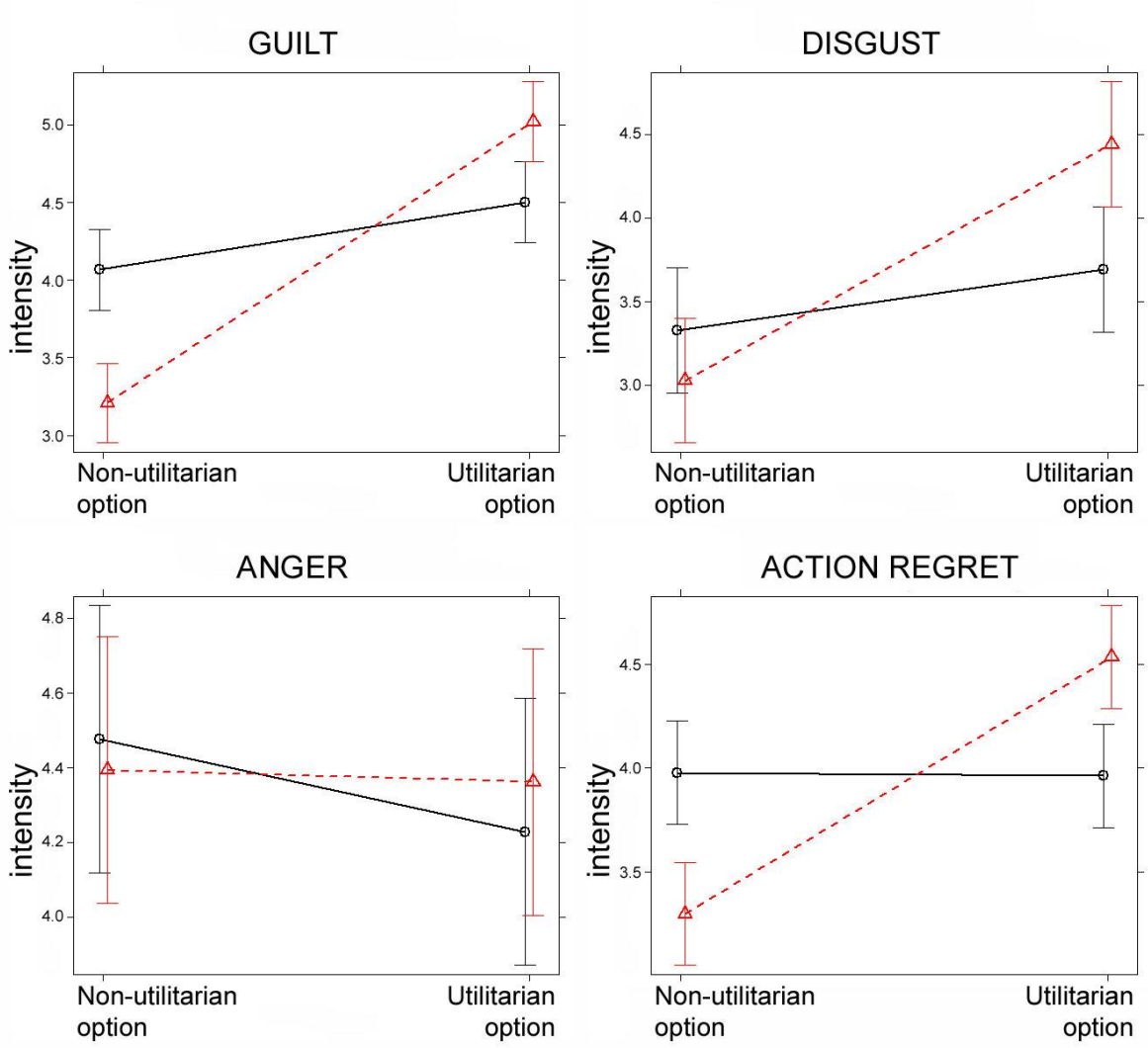
4.3. Results

4.3.1. Behavioral data

4.3.1.1. Emotional intensity ratings

For every emotion, the best model included the interaction between dilemma type and option , that was always significant (guilt: $B = 1.36$, $SE = .09$, $t(3928) = 15.13$, $p < .001$, $\chi^2(1) = 222.49$, $p < .001$; disgust: $B = 1.10$, $SE = .09$, $t(3923) = 12.25$, $p < .001$, $\chi^2(1) = 147.45$, $p < .001$; anger: $B = 0.24$, $SE = .07$, $t(3924) = 3.38$, $p < .001$, $\chi^2(1) = 11.43$, $p < .001$; action regret: $B = 1.24$, $SE = .09$, $t(3923) = 14.29$, $p < .001$, $\chi^2(1) = 199.21$, $p < .001$; inaction regret: $B = .97$, $SE = .08$, $t(3922) = 10.96$, $p < .001$, $\chi^2(1) = 118.44$, $p < .001$; shame: $B = 1.64$, $SE = .1$, $t(3925) = 16.47$, $p < .001$, $\chi^2(1) = 262.6$, $p < .001$). Follow-up analysis performed on Footbridge-type and Trolley-type dilemmas separately showed that, as concerns guilt, disgust and shame, the interaction effect was due to higher emotional intensities reported for utilitarian as compared to non-utilitarian options for both dilemma types, and to this difference being more pronounced for Footbridge-type as compared to Trolley-type dilemmas (Trolley-type, guilt: $B = .42$, $SE = .06$, $t(1948) = 7.01$, $p < .001$; Footbridge-type, guilt: $B = 1.79$, $SE = .06$, $t(1930) = 27.06$, $p < .001$; Trolley-type, disgust: $B = .38$, $SE = .06$, $t(1947) = 6.07$, $p < .001$ Footbridge-type, disgust: $B = 1.48$, $SE = .06$, $t(1928) = 23.40$, $p < .001$; Trolley-type, shame: $B = .61$, $SE = .07$, $t(1949) = 8.98$, $p < .001$ Footbridge-type, shame: $B = 2.25$, $SE = .07$, $t(1929) = 31.52$, $p < .001$; see Figure 4.2). As concerns anger, the interaction was

due to lower emotional intensities reported for utilitarian as compared to non-utilitarian options in Trolley-type dilemmas only (Trolley-type: B = -.26, SE = .05, t(1944) = -5.06, p < .001; Footbridge-type: B = -.02, SE = .05, t(1929) = -.36, p = .72; see Figure 4.2). As concerns action regret, the interaction was due to higher emotional intensities for utilitarian as compared to non-utilitarian options in Footbridge-type dilemmas only (Trolley-type: B = -.03, SE = .06, t(1945) = -.5, p = .62; Footbridge-type: B = 1.21, SE = .06, t(1925) = 19.36, p < .001; see Figure 4.2). For inaction regret, the interaction was due to lower emotional intensities for utilitarian as compared to non-utilitarian options in Trolley-type dilemmas (B = -.52, SE = .06, t(1945) = -8.86, p < .001) and to higher emotional intensities for utilitarian as compared to non-utilitarian options in Footbridge-type dilemmas (B =.45, SE = .06, t(1924) = 6.83, p < .001; see Figure 4.2).



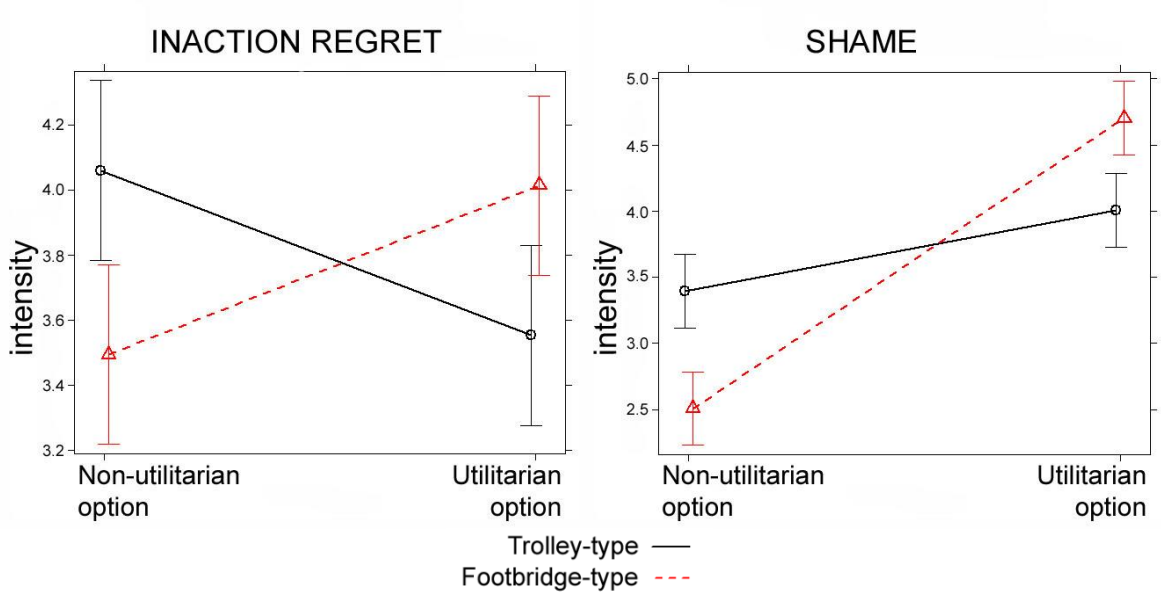


Figure 4.2. Effect of dilemma type and option on emotional intensities. The scales ranged from 0 (no intensity) to 6 (maximal intensity). Error bars indicate 95% confidence intervals.

4.3.1.2. Choices

The dilemma type effect on choices was significant (B = -2.77, SE = .26, z = -10.51, p < .001; $\chi^2(1) = 65.38$, $p < .001$): the probability of choosing the utilitarian option was higher in Trolley-type as compared with Footbridge-type dilemmas (.82 in Trolley-type dilemmas, 95% CI = [.75, .88]; .23 in Footrbridge-type dilemmas, 95%CI = [.16, .32]). As regards the effect of guilt on choices, including differential guilt intensity to the model with only the random effects significantly improved it ($\chi^2(1) = 84.31$, $p < .001$; B = -.32, SE = .04, z = -8.99, p < .001). Introducing dilemma type significantly improved the model ($\chi^2(1) = 49.52$, $p < .001$; B = -2.63, SE = .30, z = -9.06, p < .001). The differential guilt intensity effect was still significant after including dilemma type (B = -.30, SE = .04, z = -8.30, p < .001). The model with the interaction effect did not significantly differ from model with the two main effects ($\chi^2(1) = 3.79$, $p = .05$), even if the interaction term was significant (B = -.13, SD = .07, z = -2.0, p = .049).

A similar pattern of result was obtained for disgust, anger, action regret, inaction regret, shame, and for the mean emotional intensity difference: the best model was the one with both differential emotional intensity and dilemma type (disgust: $\chi^2(1) = 57.07$, $p < .001$; anger: $\chi^2(1)$ =

64.01, $p < .001$; action regret: $\chi^2(1) = 56.61$, $p < .001$; inaction regret: $\chi^2(1) = 61.49$, $p < .001$; shame: $\chi^2(1) = 52.67$, $p < .001$; mean emotional intensity difference: $\chi^2(1) = 46.47$, $p < .001$), that were both significant (differential emotional intensity: all *p*s <.001; dilemma type: all *p*s < .001), and the interaction term never improved the models (disgust: $\chi^2(1) = 1.92$, $p = .16$; anger: $\chi^2(1) = .001$, $p = .97$; action regret: $\chi^2(1) = .36$, $p = .55$; inaction regret: $\chi^2(1) = .17$, $p = .68$, shame: $\chi^2(1) = .42$, p = .52; mean emotional intensity difference: $\chi^2(1) = 2.51$, $p = .11$) nor was significant (all *p*s >.11). Thus, the difference in emotional intensities between the utilitarian and the non-utilitarian options was always negatively associated with the probability to choose the utilitarian option: the greater the emotional intensities for the utilitarian option as compared to the non-utilitarian option, the lower the probability to choose the utilitarian option, and vice versa. This effect was not influenced by dilemma type (see Figure 4.3).

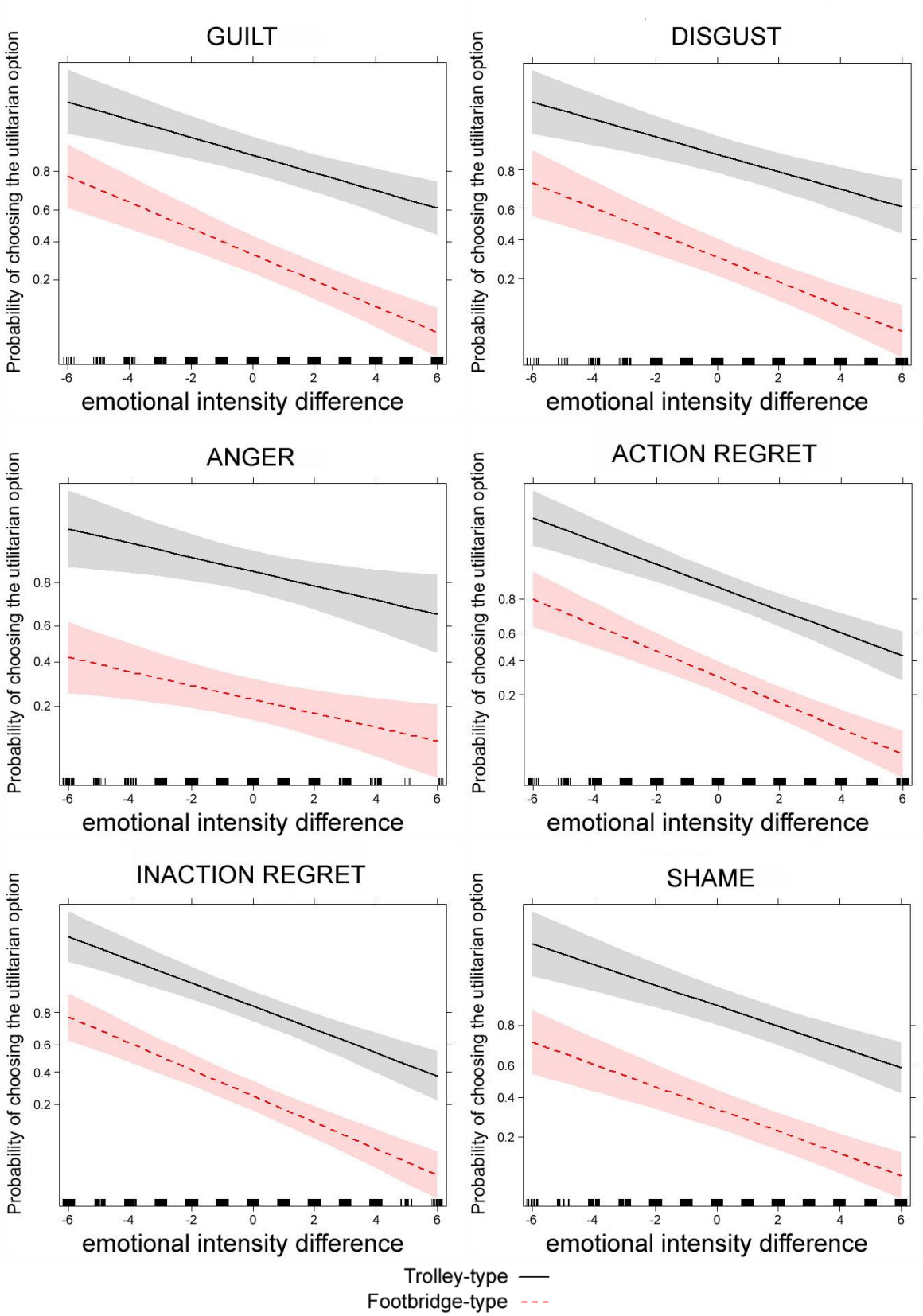


Figure 4.3. Relationship between emotional intensity difference and probability of choosing the utilitarian option, represented separately for dilemma type. Positive emotional intensity differences indicate that the utilitarian option was associated with stronger intensities than the non-utilitarian option. Negative emotional intensity differences indicate that the non-utilitarian option was associated with stronger emotional intensity differences than the utilitarian option. Shaded areas indicate 95% confidence intervals.

4.3.2. Electrophysiological data

Grand-averaged MRPs elicited before response choice in the Footbridge and Trolley groups are displayed at Cz in Figure 4.4.

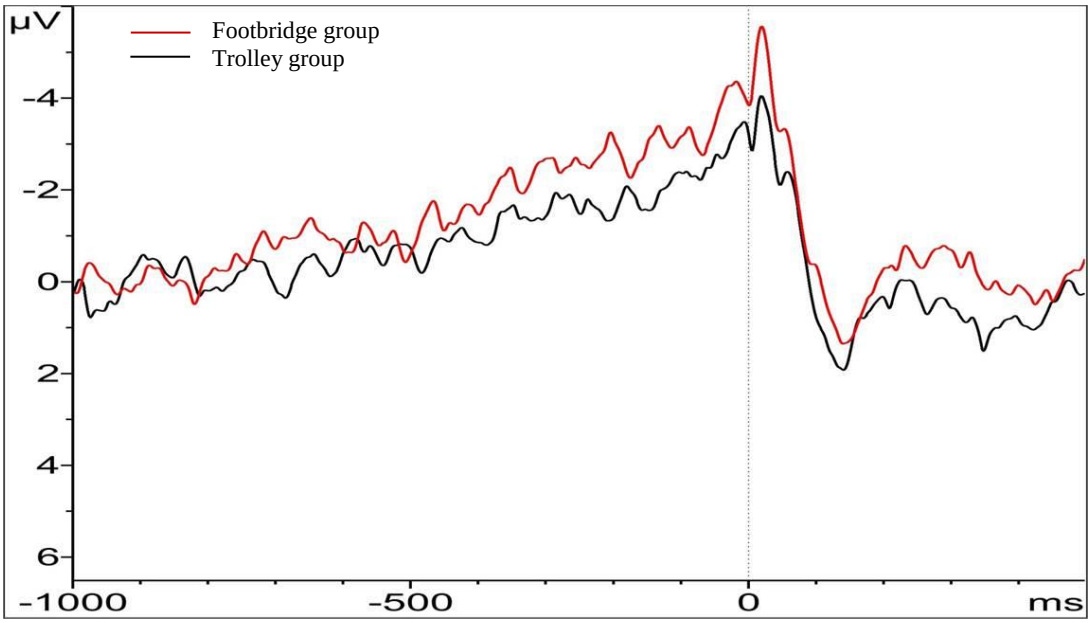


Figure 4.4. Grand-averaged MRPs recorded at representative midline sites time-locked to the behavioral response in the Trolley and Footbridge groups. Time 0 indicates the onset of the behavioral response.

4.3.2.1. Early readiness potential

The dilemma type effect was not significant: the amplitude of the early readiness potential did not differ between the Footbridge and the Trolley group ($t(40.71) = 1.17$, $p = .25$). As concerns the influence of emotional conflict on early readiness potential amplitude, none of the model was significant (all R^2s < .06, all ps > .34).

4.3.2.2. Late readiness potential

Similar result were obtained for late readiness potential. The comparison between Footbridge and Trolley groups was not significant ($t(39.35)=1.3$, $p = .02$). As for the previous time-window, none of the models that investigated the effect of emotional conflict on readiness potential amplitude was significant (all R^2s < .13, all ps > .15).

4.4. Discussion

The main aim of this study was to investigate the role of anticipated emotional consequences in driving decisions in moral dilemmas. We measured through self-report ratings the emotional state experienced by participants both after their choice and after imagining to have chosen the alternative option. The emotional state was measured on six emotions that we hypothesized to be relevant for the resolution of this kind of dilemmas: regret (translated in *rimorso* – action regret – and *rimpianto* – inaction regret), guilt, shame, anger, and disgust. Thus, for every dilemma we collected the intensity of these emotions twice, one for each option, and we analyzed it irrespective of what participants decided. We presumed that, if individuals spontaneously anticipate the emotions they would feel after the decision and use this information as input in the decision process, then the difference between the emotional states related to the utilitarian and the non-utilitarian option would predict participant's choices.

According to Greene and colleagues' dual process model (2004, 2001) utilitarian options are rejected in Footbridge-type dilemmas because they evoke strong aversive emotional reactions. Thus, we hypothesized that, in Footbridge-type dilemmas, utilitarian options would be associated with higher negative emotions as compared to non-utilitarian options. In Trolley-type dilemmas, in contrast, we anticipated the difference between emotions to be smaller.

Results on emotional intensities of guilt, shame, disgust and action regret were largely coherent with our hypothesis: the intensity of these emotions was higher for the utilitarian choices as compared to the non-utilitarian choices in Footbridge-type dilemmas, and this difference was reduced for Trolley-type dilemmas. Thus, sacrificing one person elicits more intense self-condemning emotions than letting some people die, and this difference is especially pronounced when this sacrifice is performed intentionally, as a means to an end. Sacrificing one person intentionally also elicits more *rimorso* (action regret) than letting some people die, whereas sacrificing one person as a side effect does not, coherently with an account of regret as being

strongly influenced by agency and personal responsibility (e.g., Giorgetta et al., 2012; Wagner et al., 2012; Zeelenberg, van Dijk, et al., 2000).

Results on anger and *rimpianto* (inaction regret), on the other hand, followed a different trend. Anger was stronger for the non-utilitarian option in Trolley-type dilemmas only (albeit this effect was weak), with no difference observed for Footbridge-type dilemmas. Thus, the emotion of anger is reduced by choosing the utilitarian option as compared to choosing the non-utilitarian option. This is in line with the results reported by Choe and Min (2011), who showed that high trait anger was positively associated with utilitarianism, and by Ugazio and colleagues (2012), who showed that inducing anger in participants before a moral dilemma task increased the percentage of utilitarian choices. This positive relationship between anger and the utilitarian choice could be due to the fact that anger is an approach-related emotion entailing a motivation to act, and the utilitarian choice in moral dilemmas entails action, whereas the non-utilitarian choice entails inaction.

Rimpianto (inaction regret) was stronger for the utilitarian option in Footbridge-type dilemmas, but was stronger for the non-utilitarian option in Trolley-type dilemmas. This is not surprising, given that in Italian the regret labeled *rimpianto* is less related to acting than *rimorso* (Giorgetta et al., 2012). In fact, the present study showed that in Trolley-type dilemmas this emotion was more intense for inaction than for action. This result, together with those observed for anger, indicate that choosing the utilitarian option is not only backed up by rational thinking, but also, at least in the case of Trolley-type dilemmas, by the need to avoid stronger feelings of regret and anger.

Our data are also in line with the hypothesis that individuals choose the option with the lower anticipated emotional consequences: for every emotion, the difference in emotional intensity was always significantly associated with choices, which indicates that participants chose the option with the least aversive emotional consequences. Interestingly, however, this effect was not modulated by dilemma-type: even in Trolley-type dilemmas, choices seemed to be driven by emotion, in contrast with the predictions of the dual process model, which hypothesizes that in

Trolley-type dilemmas choices are not driven by emotions, but rather by cognitive evaluations of costs and benefits. Our result do not exclude that the cognitive cost-benefit analysis plays a part in driving choices in moral dilemmas, but indicate that emotions affect decisions in both dilemma types, rather than only in Footbrige-type dilemmas.

Another important result that emerged from the present data is that the difference in choices between Trolley-type and Footbridge-type dilemmas could not be completely explained by differences in emotional consequences between options. This result seems in contrast with the dual process model, which predicts that the differences in choices between Footbridge- and Trolley-type dilemmas is due to differences in emotional intensities. On the contrary, our results show that even with emotional intensities held constant, the probability of choosing the utilitarian options was still higher in Trolley-type dilemmas than in Footbridge-type dilemmas. There are at least three possible explanations for this result: the first one is that the difference in dilemma types is produced by another emotion that we did not include in the list. For instance, we didn't measure the empathic emotions that individuals could feel for the victims. Previous literature showed that the disposition to feel empathic concern and personal distress in front of the suffering of others was inversely associated with the endorsement of the utilitarian option (Gleichgerrcht & Young, 2013; Sarlo et al., 2014). Personal distress, in particular, predicted the percentage of utilitarian choices in Footbridge-type dilemmas, but not in Trolley-type dilemmas (Sarlo et al., 2014). Thus, it is possible that the difference in the probability of choosing the utilitarian option in the two dilemma types is explained by differences in the personal distress elicited by the victims. However, as seen in Chapter 3, even individuals that have reduced empathy because of high psychopathy traits endorse more the utilitarian option in Trolley-type as compared to Footbridge-type dilemmas. Thus, at least one additional element should be introduced in the picture to fully account for the dilemma type effect on choices.

One possible candidate, as suggested by the studies of Nichols (2002) and Nichols and Mallon (2006), could be the representation of rules. According to Nichols and Mallon (2006),

individuals do not endorse the utilitarian option in Footbridge-type dilemmas because they embrace a rule that prohibits killing. This prohibition would be stronger when killing is a means to an end than when it's a side effect, because killing as a means is perceived to be more intentional than killing as a side effect. The representation of this rule against killing could concur to produce the difference in responses that is observed between Footbridge-type and Trolley-type dilemmas, together with the attempt to avoid the worst anticipated emotional consequences.

Finally, and importantly, an additional explanation could be that the difference between Footbridge-type and Trolley-type dilemmas is not due to a difference in anticipated emotions, but stems from a difference in anticipatory emotions, that are immediate visceral reactions elicited by the prospective outcomes of a choice (Loewenstein et al., 2001). Anticipatory emotions are different from anticipated emotions since the latter, as opposed to the former, are not experienced as feelings at the moment of the decision (Loewenstein et al., 2001). As described in the general introduction, according to the somatic marker hypothesis (Bechara & Damasio, 2005), the visceral reactions elicited by the possible outcomes of a choice experienced while making a decision have a strong influence on choices. Indeed, the dual process model of moral judgment, taking inspiration from the somatic marker hypothesis, proposes that, in the resolution of moral dilemmas, individuals are influenced by automatic and immediate emotional reactions that take place during the resolution of the dilemmas (Greene et al., 2004, 2001; Greene, 2008). In our study, the self-evaluations that were administered to the participants captured the hypothetical emotional consequences that participants were aware of, leaving out the visceral reactions experienced at the moment of the decision. However, such anticipatory emotions are integral part of the immediate emotions that individuals experience during the decision, and previous literature on moral dilemmas reported no direct relationship between immediate emotions and decisions. Thus, it seems unlikely that anticipatory emotions alone are the only cause of the difference between dilemma types. It is anyway possible that, during the decision process, anticipatory emotions interact with anticipated emotional consequences, for instance by making individuals focus more on anticipated emotional

consequences to the detriment of rational evaluations, or by leading participants to overestimate the intensity of anticipated post-decisional emotions during the decision (see Loewenstein et al., 2001, for a review of how anticipatory emotions influence evaluations of post-decisional consequences). Moreover, it is important to stress that self-report scales – like those that were used in the studies presented in this thesis, and by Lotto et al., 2014 and Sarlo et al., 2012 – only reflect the emotions that participants are aware of. However, affective reactions do not need to reach awareness in order to influence decisions and behaviors (e.g., Damasio, 1994). Thus, the studies presented in this thesis only captured a part of the emotional states elicited by the resolution of moral dilemmas – that is, those that participants consciously felt – and for this reason might have underestimated the effect of emotion on decisions and judgments. These are important issues that would require further investigation.

As for the RP, as opposed to previous results reported by Sarlo and colleagues (2012), in the present study we did not find a significant difference between Trolley-type and Footbridge-type dilemmas. The lack of a significant difference between dilemma types might be due to the fact that, as opposed to Sarlo and colleagues (Sarlo et al., 2012), we analyzed the RP using a between-participants design, which increased the variability of the data. Moreover and importantly, in this experiment the great part of the dilemmas that each participant solved were extracted from the same category. As a consequences, the participants of this study were overall more coherent in their choices during the task as compared to participants from Sarlo and colleagues (2012): in the study by Sarlo and colleagues (2012), participants solved both Trolley-type and Footbridge-type dilemmas in equal proportion, and thus might have been aware of the incongruence in their own choices. Conversely, in the present study, participants in the Trolley group mainly solved Trolley-type dilemmas, thus providing mainly utilitarian choices; participants in the Footbridge group mainly solved Footbridge-type dilemmas, thus providing mainly non-utilitarian choices. This coherence could have reduced the perceived conflict.

Finally, we did not found concrete evidence pointing towards a relationship between amplitude of the RP and emotional conflict. Thus, anticipated emotions did not influence the neural correlates of the last phase of decision-making, but possibly played a role in the earlier stages. This also indicates that the RP in the context of moral decisions does not reflect an emotional conflict, but rather a conflict of some other kind. Thus, the functional role of this potential in the context of moral decisions has to be further investigated.

As a main limitation of the present study, it is important to point out that in our paradigm we didn't directly measure anticipated emotions. Rather, we measured post-decisional and counterfactual emotions, and hypothesized participants to spontaneously anticipate these emotions during the decision. Some studies indicate that individuals are not always accurate in predicting how they would feel after making a choice, and that there is often a discrepancy between anticipated emotions and actual post-decisional emotions (Wilson & Gilbert, 2005). In the context of moral dilemmas, however, all the decisions that participants made are hypothetical, and participants are not confronted with real consequences. For this reason, we can expect post-decisional emotional ratings to accurately reflect the emotional consequences that participants anticipated while they were making their choices. However, it cannot be excluded that the decision itself influenced post-decisional emotional evaluations. In any case, if we asked participants to report anticipated emotions before the decision, we would have probably biased participants to take emotions into account more than they would have done spontaneously, and thus our findings on choices would have been altered. Thus, our paradigm was a good compromise as it allowed to study the role of anticipated emotions in moral dilemmas without generating considerable modifications to the decision process itself.

4.5. Conclusion

The results reported in this study indicate that in moral dilemmas participants choose the option that minimized the intensity of the emotions experienced after the decision. This effect was

present in both dilemma types and did not eliminate the differences in the probability of choosing the utilitarian option for the two dilemma types. This study thus provides useful indications for the understanding of how emotions influence the resolution of moral dilemmas. Future studies should investigate if the difference between Trolley- and Footbridge-type dilemmas that is not explained by anticipated emotions is due to the contribution of rules, to an interplay between anticipated and anticipatory emotions, or both. These issues will be further addressed in the general discussion of this thesis.

5. GENERAL DISCUSSION

The studies described in this thesis investigated the influence of emotion in the resolution of moral dilemmas in which the individual is confronted with the choice of letting some people die (non-utilitarian choice) or sacrificing a single person to save them (utilitarian choice). The aim of the studies described in this thesis was to test the dual process model of moral judgment (Greene et al., 2004, 2001), according to which emotional processing, in competition with rational reasoning, drives judgments and decisions in moral dilemmas. Moreover, the studies here described focused on how intentionality affects the emotions elicited by the dilemma on the one hand, and the choices and judgments provided by participants on the other. Furthermore, two of the three studies investigated the neural correlates of the decision process by measuring event-related potentials. The high temporal resolution of this measures allowed to separately investigate different stages of the decision, with the aim of clarifying the temporal dynamics of the decision process involved in the resolution of moral dilemmas. In particular, the study presented in Chapter 2 focused both on an earlier stage of decision-making in which the two dilemma resolutions are under evaluation and a later stage in which the action corresponding to the choice taken is implemented. The study presented in Chapter 4 focused only on the last stage, to clarify the role of the readiness potential (RP) as reflecting conflict in moral decisions.

Crucially, as opposed to the majority of the studies on moral dilemmas, which employed the original dilemma set developed by Greene and colleagues (2001), the studies presented in this thesis employed a different dilemma set, in which Footbridge-type and Trolley-type dilemmas are categorized according to the means vs side effect distinction (Lotto et al., 2014). The original set by Greene and colleagues categorizes moral dilemmas as Footbridge-type and Trolley-type according to three different criteria (see Paragraph 1.2.3 of the General Introduction), making it difficult to understand which characteristics of the dilemmas produce the difference in judgment and choices that is observed between Trolley- and Footbridge-type dilemmas, and through which mechanisms.

Conversely, the results obtained with the dilemma set devised by Lotto and colleagues (2014) can be primarily attributed to the means vs side effect distinction. Previous literature showed that the means vs side effect distinction reliably produces differences in judgments and choices even when other variables, such as the physical proximity with the victim and the personal force, are controlled (see Paragraph 1.2.3.2 of the General Introduction). Importantly, the means vs side effect distinction affects the attribution of intentionality, such that an outcome produced as a means to an end is considered more intentional than an outcome produced as a side effect (Cushman & Young, 2011). This is a relevant point because intentionality strongly influences moral judgment (Cushman, 2008; Ohtsubo, 2007; Young & Saxe, 2009b), the attribution of blame (Guglielmo et al., 2009; Ohtsubo, 2007; Shultz & Wells, 1985; Treadway et al., 2014) and punishment (Cushman, 2008; Treadway et al., 2014), and the emotional reactions to harmful actions (Treadway et al., 2014).

5.1. Summary of the research studies

The first study presented in this thesis tested the hypothesis that taking hypothetical legal consequences into account plays a role in the resolution of moral dilemmas that differ for intentionality, by bringing individuals to reject more the utilitarian option in Footbridge-type dilemmas – in which the sacrifice of one person is intentional – than in Trolley-type dilemmas – in which the sacrifice of one person is a side effect. In this study, the neural activity, subjective emotional reactions, and behavioral choices of two groups of participants were compared. One group took hypothetical legal consequences into account when deciding; the other was told that none of the options described in the dilemmas was legally prosecutable, and thus decided without considering potential legal consequences. Stimulus- and response-locked ERPs were measured to investigate the neural activity underlying the earlier stage of the decision process, in which the two options (utilitarian and non-utilitarian) were under evaluation, and the final stage, in which the action corresponding to the chosen option was performed. The results that emerged from this study showed that excluding legal consequences from the decision-making process did not influence the

effect of intentionality on choices, but did modulate the neural correlates of the decision and its emotional impact. In particular, the group that did not take legal consequences into account reported greater overall emotional engagement during decision-making, associated with lower preparation for action, suggesting a greater conflict between alternative motor responses representing the different decision choices. In contrast, the group that did take legal consequences into account showed an overall dampened affective experience during decision-making associated with greater overall action readiness and intention to act, reflecting lower conflict in responding. These results were interpreted as indicating that in moral dilemmas legal consequences of actions provide a sort of reference point on which people can rely to support decision-making, independent of dilemma type. The results emerged in this study were generally in line with what hypothesized by the dual process model, as Footbridge-type dilemmas produced less utilitarian options and a more unpleasant emotional experience during their resolution as compared to Footbridge-type dilemmas. However, no significant correlation emerged between choices and self-reported emotional experience.

The second study aimed at investigating the effect of psychopathy – a personality trait characterized by emotional hyporeactivity to harm and by a propensity to immoral behavior (Blair, 2011; Cleckley, 1976; Hare, 2003) – on choices, judgments and on the emotional state experienced during decision-making in dilemmas that differ for intentionality. A group of participants with high trait psychopathy was compared with a group of participants with low trait psychopathy. We hypothesized participants with high trait psychopathy to experience a less negative emotional state during the decision, associated with a higher endorsement of the utilitarian resolutions for both moral judgment and choices. Moreover, we expected this effect to be specific for Footbridge-type dilemmas, since, according to the dual process model of moral judgment (Greene et al., 2004, 2001), Trolley-type dilemmas would generally not elicit a strong emotional reaction and their resolution would be driven by a cognitive cost-benefit analysis even in individuals wo do not present emotional hyporeactivity.

The results only partially supported the hypotheses: in line with the hypothesis, participants with high trait psychopathy experienced lower unpleasantness during the resolution of moral dilemmas and were more likely to choose to sacrifice one person to save others, as compared to participants with low trait psychopathy. However, as opposed to what hypothesized, these effects were not specific for Footbridge-type dilemmas, but were comparable for both dilemma types. Moreover, no significant correlations between self-reported emotions and decision choices were found. Furthermore, the difference between choices and judgments in Trolley-type and Footbridge-type dilemmas had the same direction and magnitude in the two groups: both participants with low trait psychopathy and participants with high trait psychopathy showed a lower endorsement of the utilitarian option in Footbridge- as compared to Trolley-type dilemmas. This suggests that it may not be necessary to feel a stronger emotional reaction in order to reject the utilitarian option in Footbridge-type dilemmas more than in Trolley-type dilemmas, which also implies that another mechanism apart from emotional processing might be involved in producing the differences in choices and judgments that is observed for Footbridge-type and Trolley-type dilemmas. Coherently with this accout, no difference emerged between participants with high and low psychopathy traits in moral judgments. This last result also suggests that emotional processing might play a greater role in choices than in judgments, because trait psychopathy – which is characterized by emotional hyporeactivity – influenced both the emotional state experienced by participants during the decision and the inclination to endorse the utilitarian choice, but did not affect moral judgments.

Thus, the results presented in this study were partially in line with the dual process model (Greene et al., 2004, 2001), since they showed that emotional hyporeactivity – as characteristic of individuals with high psychopathy traits – produces not only a less unpleasant emotional state during the resolution of the dilemmas, but also a higher endorsement of the utilitarian option (at least as regards choices). However, these results also suggests that emotional processing plays a role both in Footbridge-type and in Trolley-type dilemmas, as opposed to what hypothesized by the model, since the effect of psychopathy traits on choices and affective ratings emerged for both

dilemma types. Moreover, as in the previous study, the association between choices and emotional states experienced during decision-making was not significant.

To further investigate the relationship between emotion and decisions-making in moral dilemmas, the last study focused on the role of anticipated emotions in driving individuals to reject the utilitarian resolution in the dilemmas. The anticipation of post-decisional emotions has a great impact in decisions, since individuals choose the option that is associated with less anticipated negative emotions (Bell, 1982, 1985; Loomes & Sugden, 1982, 1986; Mellers et al., 1999). We hypothesized the rejection of the utilitarian option in Footbridge-type dilemmas to be driven by the fact that people expect to feel worst if they choose to intentionally sacrifice one person than if they choose to let some people die. In this last experiment, we also tested whether the neural activity related to the last phase of the decisional process – that is, the implementation of the chosen option – was related to an emotional conflict – in terms of anticipated emotions –between the utilitarian and the non-utilitarian resolution. To this aim, we analyzed post-decisional emotions elicited by both options irrespective of the choice taken, and we calculated an index of emotional conflict between options by measuring the difference between these two emotional states. Results showed that participants indeed chose the option that minimized negative emotions, since this index of differential emotional intensity predicted the probability of choosing the utilitarian choice: the higher the emotional intensity for the utilitarian option as compared to the non-utilitarian one, the lower the probability of choosing the utilitarian option, and vice versa. Interestingly, this effect was comparable for both dilemma-types: in both Footbridge and Trolley-type dilemmas, differential emotional itnensity predicted choice. Moreover, keeping emotion constant, the probability of choosing the utilitarian option was still higher for Trolley- as compared to Footbridge-type dilemmas. This suggests that anticipated emotions are a driving factor in moral decisions, but it is likely that they are not the only reason behind the difference between Footbridge-type and Trolley-type dilemmas. The results emerged in this study also provided novel information on the emotions elicited by the utilitarian and the non-utilitarian option for the two dilemma types: In Footbridge-

type dilemmas, the utilitarian option was associated with stronger guilt, action regret, inaction regret, disgust and shame as compared to the non-utilitarian option. In Trolley-type dilemmas, the utilitarian option was associated with stronger guilt, disgust and shame as compared to the non-utilitarian option (albeit these differences were smaller if compared to those emerged for Footbridge-type dilemmas), but the non-utilitarian option was associated with stronger inaction regret and anger as compared to the utilitarian option. This indicates that the endorsement of the non-utilitarian option is not exclusively due to a "cold" cost-benefit analysis, but it is also influenced by emotions, inaction regret and anger in particular. As concerns the neural correlates of the last phase decision-making, we found no relationship between emotional conflict and cortical activity in this phase of the decision, which suggests that anticipated emotions might play a role in earlier stages of the decision.

Thus, also this final study suggests that emotions influence choices in moral dilemmas, coherently with what the dual process model (Greene et al., 2004, 2001) hypothesized, by finding direct associations between choices and the emotions reported by participants relative to the two options. However, as opposed to what hypothesized by the dual process model, the results emerged in this final study suggest that emotions have the same weight in driving choices for both dilemma types (as least as far as concerns anticipated emotions) and that the difference in choices between Footbridge-type and Trolley-type dilemmas might not be exclusively due to differences in the emotions elicited by the options presented in the dilemmas.

Taken together, the studies presented in this thesis are in line with the dual process model in indicating that emotions affect moral dilemmas resolution, but they contrast with the model's prediction by suggesting that emotions affect Trolley- and Footbridge-type dilemmas in a comparable way, and by suggesting that decision-making in moral dilemmas might not only involve emotional processing and cos-benefit analysis, but also one additional process that concurs in producing the difference in choices and judgments between the two dilemma types. This point will be further address in Paragraph 5.3.

5.2. Limitations of current research

The research studies presented in this thesis share some important limitations with other related studies in the literature. One limitation regards the use of self-report measures to assess emotional processes, and the other is a more general limitation concerning the use of moral dilemmas in studying moral cognition.

5.2.1. The measurement of emotional processes

In the studies presented in this thesis, the emotional processes involved in decision-making were measured through self-report scales, which rely on individuals' ability to monitor and assess their emotional state. However, individuals are not always aware of their emotional states, or accurate in assessing them. Thus, self-report measures provide only an approximation of the real emotions felt by participants (Larsen & Fredrickson, 1999). As concerns the results presented in this thesis, it is especially important to consider that they only provide information about those emotional states that reached awareness. In fact, it is well acknowledged that emotional and somatic reactions do not need to reach awareness in order to influence decisions and behaviors (see, for instance, Damasio, 1994). Thus, it might be the case that the role played by unconscious emotional processes in the resolution of moral dilemmas is greater (or different) than what emerged in the present research studies.

One way to capture emotional states that does not rely on individuals' awareness is through methods such as psychophysiological indexes, or through brain imaging methods. However, it is worth noting that also measures of physiological arousal or observation of brain activity in emotion-related areas only capture one part of the processes at play during an emotional experience. To have a more precise picture of what individuals feel, such measures should be associated with subjective state measures, and the relationship between these two aspects should be assessed. This is especially relevant as psychophysiological indexes such as the SCR and the heart rate, for instance, respond not only to emotion, but also to cognitive effort or to changes in attentional states (see, for instance,

Larsen & Fredrickson, 1999). Furthermore, as regards brain activity, only very few brain areas have such a high specificity for emotion that their activation can be solely attributed to it (see, for instance, Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Thus, the accurate assessment of the emotional processes involved in decision-making is an ambitious goal that will perhaps never be fully achieved (see Larsen & Fredrickson, 1999 for a discussion about the measurement of emotions that includes the issues mentioned above). Future studies aimed at investigating the role of emotions in moral dilemmas should consider a more global approach including simultaneous self-report and psychophysiological measures, striving to include as many measures as possible without compromising participants' performance.

5.2.2. Is research on moral dilemmas really impactful?

Research on moral dilemmas has been often criticized: it has been argued, for instance, that such extreme and abstract situations are far away from the moral problems that people encounter in everyday life, and thus are not a good paradigm for studying moral cognition (see, for instance, Bauman, McGraw, Bartels, & Warren, 2014).

However, although it is true that we are seldom directly involved in situations similar to those depicted in moral dilemmas, it is worth noting that we are often confronted with similar problems through newspapers and TV news. For instance, the war conflict that European nations are currently facing has often posed to our governments the choice between paying a ransom to save a few captives from terrorists – which finances terrorisms and indirectly encourages the kidnapping of other people – or letting those people die, thus reducing the effectiveness of the kidnappings and probably its occurrence. When we are confronted with such events through the news, we find ourselves imagining what we would have done in such situations and we make moral judgments. Those judgments are indeed relevant because they influence the public opinion on governments, law enforcers or armies that acted in those dilemmatic situations, and the public opinion ultimately has a great impact in modifying society. Thus, studying moral dilemmas may not be informative on

the single individual's moral choices, but it provides information on large-scale moral decisions and in moral behaviors at a societal level.

Another recent practical application of research in moral dilemmas regards automatic vehicles. Recent advances in technology allowed the implementation of automatic, driver-less, cars. Their programmers are now faced with the issue of defining algorithms that guide those car's behavior in emergency situations in which the passenger's life is at stake against those of other people: for instance, should the car run over a group of pedestrians or steer out of the road, thus harming its passengers? Researchers are attempting to program the cars so that their "moral" behavior would not cause outrage in the public. This is necessary both to avoid legal issues to the producers and to give a positive image to prospective buyers (Bonnefon, Shariff, & Rahwan, 2015). Similar problems are posed with other kinds of robots, so that machine morality is now a developing research topic in robot design (see, for instance, Malle, 2015).

5.3. Mechanisms behind decisions and judgments in moral dilemmas

Drawing from the existing literature and from the findings emerged from this thesis, we can attempt to build a sketch of the mechanisms and processes at play during the resolution of moral dilemmas, and of how they interact, with the aim of guiding future research toward the aspects that are still in need of clarification.

We can hypothesize individuals faced with moral dilemmas to assess both the outcomes of the two resolutions and the presence of harmful intentions behind the action depicted in the utilitarian resolution. These aspects would influence different interacting processes, which would provide "value weights" to the two options. We can hypothesize these processes to be at least four:

- The presence and intensity of anticipatory emotions;
- The anticipation of post-decisional emotion;
- A cost-benefit analysis made on the number of lives to save or sacrifice;
- The representation of norms, with consequent representation of punishment.

The influential power that these processes *per se* have on decision-making would vary depending both on individual differences and contextual aspects. For instance, the cost-benefit analysis would have more weight in individuals who prefer a rational problem-solving style (Bartels, 2008; Paxton, Ungar, & Greene, 2012; Wiech et al., 2013), or in contexts that encourage cold reasoning (Aguilar, Brussino, & Fernández-Dols, 2013; Bartels, 2008; Paxton et al., 2012), whereas anticipatory and anticipated emotions would have less weight in the choices of individuals who habitually resort to reappraisal strategies (Szekely & Miu, 2015), and more weight in the choices of individuals with a high inclination to experience emotions such as empathy (Choe & Min, 2011; Gleichgerrcht & Young, 2013; Sarlo et al., 2014), disgust (Choe & Min, 2011), anger (Choe & Min, 2011), or in contexts in which the emotional impact of the choice outcomes is made more salient (an aspect that has not been tested yet by the existing literature on moral dilemmas).

Drawing from recent decision-making studies and models (e.g., Bogacz, 2007; Gluth, Rieskamp, & Buchel, 2012; Grabenhorst & Rolls, 2011), we can hypothesize these processes to work in parallel, providing weight in favor of both options until the preference for one option over the other reached a threshold. This threshold would adjust depending on the motivation that individuals have to decide and on the time available for the decision: the higher the motivation and the lower the time, the lower the threshold. Moreover, we can also hypothesize these processes to be characterized by different activation speeds: anticipatory emotions, for instance, would arise early and automatically in response to specific dilemma features (cf. Sarlo et al., 2012), whereas anticipated emotions might be slower because they would require individuals to actively imagine how they would feel after the decision.

The influence of the assessment of outcomes and intentionality on these four processes, and the interrelationships between them, will be described in the next paragraphs.

5.3.1. Anticipatory emotions

The term anticipatory emotions indicates the visceral reactions elicited during the decision-making process by prefiguring the possible choice outcomes (Loewenstein et al., 2001), a concept that is akin to that of somatic markers (Bechara & Damasio, 2005). According to the somatic marker hypothesis, prospecting post-decisional consequences triggers these somatic states through the activation of the ventromedial prefrontal cortex (vmPFC) (Bechara & Damasio, 2005)[3]. These somatic state signals would help to endorse or reject the choice options that are represented in working memory during the decision-making process.

We can assume such powerful and unconscious emotional reaction to correspond to the alarm-bell emotion that, according to the dual process model (Greene et al., 2004, 2001; Greene, 2008), would lead to the rejection of the utilitarian option in Footbridge-type dilemmas. In moral dilemmas, anticipatory emotions would be influenced not only by the harmfulness of the outcome *per se*, but also by the intentionality with which the outcome is obtained, so that intentionally sacrificing a person would elicit stronger anticipatory emotions than sacrificing a person as a side effect. As described in Paragraph 1.2.4.3 of the General Introduction, the appraisal of intentionality of an action modulates the responsiveness of the amygdala to the harmful outcome of the action (Treadway et al., 2014). We can hypothesize a similar mechanism to be at play with anticipatory emotions in moral dilemmas, with the appraisal of intentionality in the temporoparietal junction (TPJ) down-regulating the vmPFC circuit subtending anticipatory emotions. Indeed, the connectivity between the TPJ and the vmPFC increases when individuals are faced with choices entailing norm violations in social contexts (Makwana, Gron, Fehr, & Hare, 2015). A goal for future studies would be to test whether solving moral dilemmas also entails enhanced connectivity

[3] It is important to note that somatic markers can also be simulated in the brain through the as-if body loop, being thus able to affect decision-making even without eliciting a full-fledged somatic response (Bechara & Damasio, 2005).

between these areas, and whether this is related to the influence of intentionality on the visceral reactions experienced during the decision.

An alternative, and maybe more parsimonious, account would be that intentionality influences anticipatory emotions indirectly, by enhancing the vividness with which the outcome of an action is represented, since vividness strongly influences the intensity of anticipatory emotions (see Loewenstein et al., 2001, for a review). Thus, the effect of intentionality on anticipatory emotions might be explained by hypothesizing intentional harm to generate more vivid representations of the victim than unintentional harm. Coherently with this hypothesis, a recent study reported that Footbridge-type dilemmas generate more vivid visual representations of the outcomes than Trolley-type dilemmas (Amit & Greene, 2012). However, these results were found on moral dilemmas categorized as personal and impersonal based on the distinction proposed by Greene and colleagues (2001), and thus it has still to be confirmed if the means vs side effect distinction would produce the same results.

5.3.1. Anticipated emotions

Anticipated emotions, also labeled as "expected emotions" (Loewenstein & Lerner, 2003) are emotional states that individuals expect to feel as a consequence of the outcome of a decision.

As seen in Chapter 4, the anticipation of post-decisional emotion plays an important role in the resolution of moral dilemmas, as participant attempt to choose the option associated with the least intense negative post-decisional emotions. The results of the study presented in Chapter 4 showed how post-decisional emotions are influenced both by the outcomes and the intentionality of the action depicted in the dilemmas: the utilitarian option elicited overall stronger self-condemning emotions than the non-utilitarian option, with this difference being larger when the sacrifice of one person was performed intentionally. However, when the sacrifice of one person was a side effect, the non-utilitarian option generated stronger emotions of anger and regret. The fact that the difference in the intensity of post-decisional emotions elicited by the two options predicted the

probability of choosing the utilitarian option suggested that, during the resolution of moral dilemmas, individuals anticipated how they would have felt after the decision and then chose the option associated with the least intense negative emotions.

Other anticipated emotions that could be relevant for moral dilemmas, but were not investigated in the present research, are the empathic emotions elicited by the victims. The literature showed that a high disposition to experience affective empathy is inversely related with the endorsement of the utilitarian option (Gleichgerrcht & Young, 2013; Sarlo et al., 2014), and that this effect is modulated by intentionality, being stronger for Footbridge- than for Trolley-type dilemmas (Sarlo et al., 2014). Thus, we can hypothesize that anticipating empathic emotions for the victim would motivate participants to reject the choice of sacrificing one person, especially when the sacrifice is to be performed intentionally.

Anticipated emotions can be hypothesized to interact with anticipatory emotions in a bidirectional way. On the one hand, anticipatory emotions might amplify the influence of anticipated emotions on decisions by making individuals focus more on them, or by making individuals overestimate their intensity (see Baumeister, Vohs, Nathan DeWall, & Liqing Zhang, 2007; Loewenstein & Lerner, 2003, for reviews). On the other hand, the anticipation of post-decisional emotions might *per se* generate anticipatory emotions (Loewenstein & Lerner, 2003), which in turn might affect decision-making. It is important to note, however, that the debate is still open as to whether the anticipatory emotions felt during the decision are weaker versions of the emotion that would be experienced after the decision, for instance guilt or regret (Baumeister et al., 2007; Chang et al., 2011; Coricelli et al., 2005), completely different emotions, like fear or anxiety (Loewenstein & Lerner, 2003; Loewenstein et al., 2001), or both. Moreover, it is difficult to distinguish anticipated from anticipatory emotions at the neural level: the vmPFC, for instance, is not only crucial for triggering anticipatory emotions (Bechara & Damasio, 2005), but it is also involved in the anticipation of post-decisional regret (Coricelli et al., 2005). The relationship between anticipatory and anticipated emotions would thus need further research to be fully

understood. A better comprehension of the interplay between these two processes would then allow a further understanding of their influence in decision-making.

5.3.2. Cost-benefit cognitive analysis

One of the processes at play during the resolution of moral dilemmas would be a cost benefit analysis focused on the physical outcomes of the choice (i.e., the number of deaths, or lives) that would always favor the option that minimizes the number of deaths (or maximize the number of lives). This computation would have a greater weight for people who are inclined to solve problems through rational reasoning and logic (Bartels, 2008; Paxton et al., 2012; Wiech et al., 2013), and would acquire more value in decisional contexts that encourage cold deliberation (Aguilar et al., 2013; Bartels, 2008; Paxton et al., 2012).

The cost-benefit analysis on the number of deaths appears to have a greater weight in Trolley- than in Footbridge-type dilemmas, since in the former dilemmas choices and judgments are more often coherent with this evaluation than in the latter (see, for instance, Greene et al., 2004, 2001, and the results presented in this thesis). One aspect that still requires investigation is whether this pattern reflects an inhibitory effect of the assessment of intentionality on this type of evaluation –that is, the cost-benefit analysis is not performed at all, or assumes a lower weight, when the sacrifice of one person is intentional – or simply the fact that in Footbridge-type dilemmas the non-utilitarian option outweighs the utilitarian one because of the weights that the other processes at play assign to it – that is, the cost-benefit analysis is performed irrespective of the intentionality of the action and would lead to choosing the utilitarian resolution, but the other processes at play would jointly lead to a stronger preference for the non-utilitarian resolution when the action is intentional.

The information provided by the literature about which neural mechanisms would implement this cost-benefit analysis is not clear up to now: the first studies by Greene and colleagues (2004, 2001) suggested a role of brain areas related to working-memory and cognitive

control, like the dorsolateral prefrontal cortex (dlPFC) and the inferior parietal lobe, but more recent findings by Hutcherson and colleagues (2015) found utilitarian evaluations to be associated to activity in areas related to social cognition like the TPJ (that is also involved in the assessment of intentionality) and the dorsomedial prefrontal cortex (dmPFC). Further research is thus needed to shed light on this process and its neural bases.

5.3.3. Representation of norms

The fact that individuals with high trait psychopathy provide moral judgments that do not differ from those of people with low trait psychopathy (see Chapter 3 of this thesis) suggests that the rejection of the utilitarian choice might not be motivated only by emotional processing: even without a stronger emotional reaction to intentional harm, individuals with high psychopathy traits are still able to judge the utilitarian option as less morally acceptable in Footbridge- than Trolley-type dilemmas. Similar results have been reported by several previous studies, which attributed this effect to the fact that these participants rely on abstract reasoning and on the knowledge of moral prohibitions when they provide moral judgments to moral dilemmas (Glenn, Raine, Schug, et al., 2009; Tassy, Deruelle, et al., 2013). This implies the existence of a rule that individuals refer to during the resolution of moral dilemmas. This rule would be violated in Footbridge-type dilemmas, but not (or less severely) in Trolley-type dilemmas. According to authors like Nichols and Mallon (2006), it would be a rule against killing, whose violation would be more severe in Footbridge-type dilemmas, since in these dilemmas the killing is carried out intentionally. These authors proposed that individual's moral judgments are based not only on the emotional reactions elicited by the outcomes of an action, but also on a *normative theory* – that is, a body of norms describing what is allowed and what is not (Nichols, 2002). This normative theory would be acquired through learning during the development: social norms are explicitly taught to children by adults, who also sanction behaviors that are not compliant with the rules; norms are extracted by children through observation

of social interactions in different contexts; finally, norms are consolidated through role-playing with peers (Buckholtz & Marois, 2012; Darley & Shultz, 1990).

According to Nichols and Mallon (2006), the mere presence of a rule is not sufficient to lead individuals to judge as morally impermissible an action that has favorable outcomes: in the absence of emotionally aversive outcomes, a cost-benefit analysis would prevail and would bring participants to judge the action as permissible despite the rule. On the other hand, emotionally aversive outcomes are not sufficient to bring individuals to judge as morally impermissible an action, in the absence of proscriptive rules against it. For instance, acts of self-defense or violent acts carried out during wars against oppressor are not typically judged as morally impermissible. Moreover, the authors point out that there are cultures in which it is not considered wrong for men to beat their wives – despite it being a violent and emotionally aversive act – because there is no rule against it. Even in our culture, less than a century ago, the corporal punishment of children by teachers was not considered morally unacceptable, despite it being violent and emotionally aversive. Thus, judging an action as morally impermissible would require both the presence of emotionally aversive outcomes and the presence of a norm proscribing the action.

We can speculate the processes involved in norm-compliant behavior to be based on the dlPFC, since this area is more active when participants with high psychopathy traits provide moral judgments that are analogous to those provided by participants with low trait psychopathy (Glenn, Raine, Schug, et al., 2009). Moreover, interfering with dlPFC functioning through transcranial magnetic stimulation (TMS) produces an increase in utilitarian judgments in moral dilemmas (Tassy et al., 2012).

In moral dilemmas, the representation of rules would influence not only judgments, but also choices: participants would prefer to avoid breaking a rule in order to maintain a positive self-concept (Mazar, Amir, & Ariely, 2008) and to avoid blame and punishment (Buckholtz & Marois, 2012).

As concerns the interaction between rule representation and other processes involved in the resolution of moral dilemmas, we can hypothesize the appraisal of a rule violation to influence both anticipatory and anticipated emotions: on the one hand, the appraisal of a rule violation might cause anticipatory fear related to the idea of being punished; on the other hand, it might generate greater anticipatory emotions such as regret and – especially – shame, which is strongly related to the violation of social rules (Tangney et al., 1996). However, as seen in Chapter 2, taking legal consideration into account – that is, explicit and externally enforced rules – also reduces the conflict inherent to the decision and its unpleasantness, possibly because legal considerations serve as a reference point.

5.3.4. The neural substrates of decision-making and judgments in moral dilemmas

When we attempt to sketch the neural bases of the processes involved in the resolution of moral dilemmas that were identified in the previous paragraphs, we find that the same brain areas are involved in almost all of the aspects. For instance, the vmPFC/OFC would be involved in triggering anticipatory emotions (Bechara & Damasio, 2005), in anticipating post-decisional regret (Coricelli et al., 2005), and in computing an integrative value weight that collect information from different types of appraisals (Hutcherson et al., 2015; Shenhav & Greene, 2014); the insula would be involved in the perception of anticipatory emotions (Bechara & Damasio, 2005) and in anticipating post-decisional guilt (Chang et al., 2011); the dlPFC would be involved in cost-benefit analysis (Greene et al., 2004, 2001) and in rule-compliant behavior (Buckholtz & Marois, 2012). Several authors have attempted to clarify the brain structures involved in moral judgment and moral behavior, but none of the proposals successfully accounts for all the findings emerged in the literature.

A first descriptive account proposed by Greene and Haidt in 2002 (see Figure 5.1) identified the areas involved in moral judgments as the medial frontal gyrus (BA 9/10); the posterior cingulate cortex (PCC), precuneus and retrosplenial cortex (BA 31/7); the superior temporal sulcus (STS) and

the inferior parietal lobe (BA 39); the OFC/cmPFC (BA 10/11); the temporal pole (BA 38); the amygdala; the dlPFC (BA 9/10/46) and the parietal lobe (BA 7/40). The BA 9/10 would be involved in the integration of emotions into decision-making and in theory of mind, the BA 31/7 in imaging complex situations and in retrieving emotional information from memory; the BA 39 in theory of mind and in representing socially significant information; the BA 10/11 and the amygdala in the representation of reward and punishment; the BA 39 in theory of mind; and the BA 9/10/46 and BA 7/40 in working memory and other cognitive functions not further specified by the authors.

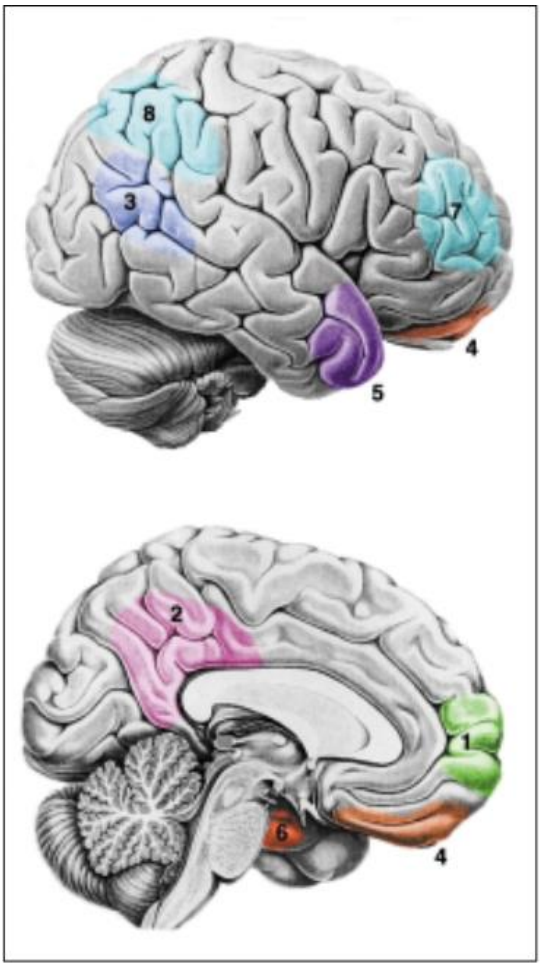


Figure 5.1. Brain areas implicated in moral cognition identified by Greene and Haidt (2002) (Brodmann's areas in parentheses): 1. medial frontal gyrus (9/10); 2. posterior cingulate, precuneus, retrosplenial cortex (31/7); 3. superior temporal sulcus, inferior parietal lobe (39); 4. orbitofrontal, ventromedial frontal cortex (10/11); 5. temporal pole (38); 6. Amygdala; 7. dorsolateral prefrontal cortex (9/10/46); 8. parietal lobe (7/40).

A second descriptive account, mainly drawing from lesion studies, was proposed by Moll, de Oliveira-Souza and Eslinger in 2003 (see Figure 5.2). According to these authors, the areas involved in moral behavior would be the medial orbitofrontal cortex, which would orchestrate automatic social-emotional responses; the frontopolar cortex, which would be responsible for emotional self-regulation and planning; the anterior cingulate cortex (ACC), which would be involved in conflict detection and emotional control; the STS, involved in the processing of socially relevant information, such as the appraisal of intentionality; the anterior temporal cortex, amygdala, insula and precuneus, which would be part of a network that integrates social and self-related emotions into perceptions and ideations. Finally, the thalamus, the midbrain and the basal forebrain were included by the authors in the model since lesions in these areas produce deviant and inadequate social behavior, which is intuitively seen as immoral.

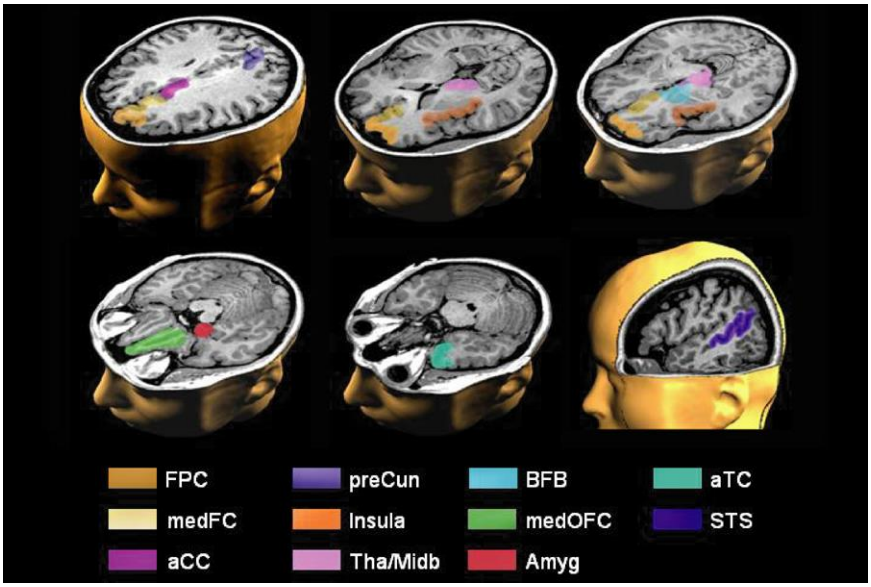


Figure 5.2. The moral brain according to Moll and colleagues (2003). FPC: frontopolar cortex, preCun: precuneus; BFB: basal forebrain; aTC; anterior temporal cortex; medFC: medial frontal cortex; medOFC: medial orbitofrontal cortex; STS: superior temporal sulcus; aCC: anterior cingulate cortex; Tha/Midb: thalamus/midbrain; amyg: amydgala.

Finally, Buckholtz and Marois (2012) defined a framework for the appraisal of third-party moral violation and the attribution of retributive punishment. According to their model, the TPJ would be responsible for encoding information about intent, which are essential for the attribution

of blame and punishment; the amygdala is involved in the affective reactions to the harmful outcomes of an action, which would influence the severity of the punishment; the medial prefrontal cortex would integrate information arising from the amygdala and the TPJ and transmit them to the dlPFC; the intraparietal sulcus would represent contextual information; finally, the dlPFC would integrate all these information and select a final punishment response option. This proposals attempts also at defining the connections of these brain areas, which would be mainly bidirectional (see Figure 5.3).

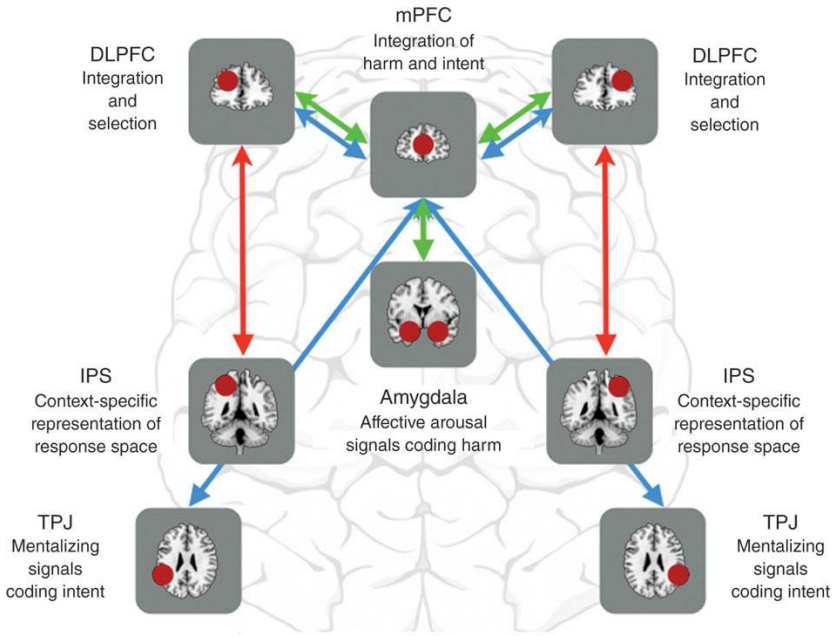


Figure 5.3. Graphical representation of the model proposed by Buckholtz and Marois (2012) on retributivist punishment. DLPFC: dorsolateral prefrontal cortex; mPFC: medial prefrontal cortex; IPS: intraparietal sulcus; TPJ; temporoparietal junction. Bidirectional arrows emphasize the interactive computations likely to take place along the core components of this network. Long red arrows, scale representation information from IPS; blue arrows, mentalizing output from TPJ; short green arrows, affective arousal signals from the amygdala.

What emerges from the comparison of these models is a rather high consistency in identifying which areas are involved in moral behavior, in moral judgment and in the attribution of punishment. However, the exact functions of these areas and, especially, their interrelationships are still unclear.

It is worth noting that the proposals described above are based on reviews of the literature that, in identifying the brain areas involved in moral cognition, rely on the anatomical labeling of

brain areas reported in the papers. This is a potentially critical aspect, because the labelling of some brain areas is not univocal and it is not always consistent in the literature. For instance, the labels OFC and vmPFC are often used as synonyms, but they refer to areas that are only partially overlapping. The labels TPJ, inferior parietal lobule, and STS are subject to a similar problem, since the three terms define partially overlapping regions. As a result, different papers may use the same label to refer to different brain regions, or different labels to refer to the same regions. This issue is exacerbated by the fact that the brain structures involved in higher cognitive functions present a great intra-subject variability, so that it is often difficult to choose reliable labels (Brett, Johnsrude, & Owen, 2002). To overcome this issue, research on moral dilemmas would greatly benefit from a coordinate-based (Wager, Lindquist, Nichols, Kober, & Van Snellenberg, 2009) or image-based (Salimi-Khorshidi, Smith, Keltner, Wager, & Nichols, 2009) meta-analysis on the fMRI studies conducted on this topic.

5.4. Conclusions

Within the last decade, a growing amount of research in psychology and cognitive neuroscience has been devoted to exploring the causes, mechanisms and aims of moral cognition. This growing interest is not surprising, given that it is through moral principles, laws, and norms that our increasingly complex societies maintain stability. On the other hand, cultural differences in moral principles are also detrimental for societal wellbeing, because they can cause the rise of xenophobia, violence and, ultimately, wars. Thus, understanding what makes people classify actions as "right" or "wrong", why humans experience moral outrage, shame or guilt, and what causes and influences moral judgments and decisions is of greatest importance for understanding both individuals and societies.

What emerges from the literature and from the studies presented in this thesis is that moral decision is a complex phenomenon involving the interplay and interaction of different processes, including emotional processing, rule representation, and cost-benefit analysis. However, several

important issues are far from settled and require further investigation. Future research should thus strive to overcome the limits that the current literature presents and to increase our understanding of the processes involved in moral decisions and their interplay.

6. REFERENCES

Aguilar, P., Brussino, S., & Fernández-Dols, J.-M. (2013). Psychological distance increases uncompromising consequentialism. *Journal of Experimental Social Psychology*, *49*(3), 449–452. doi:10.1016/j.jesp.2013.01.002

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, *23*(8), 861–868. doi:10.1177/0956797611434965

Barrett, L. F., & Fossum, T. (2001). Mental representations of affect knowledge. *Cognition & Emotion*, *15*(3), 333–363. doi:10.1080/02699930125711

Barry, C. T., Frick, P. J., DeShazo, T. M., McCoy, M., Ellis, M., & Loney, B. R. (2000). The importance of callous-unemotional traits for extending the concept of psychopathy to children. *Journal of Abnormal Psychology*, *109*(2), 335–340. doi:10.1037//0021-843X.109.2.335

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*(2), 381–417. doi:10.1016/j.cognition.2008.03.001

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*(1), 154–161. doi:10.1016/j.cognition.2011.05.010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4, 51. Computation. Retrieved from http://arxiv.org/abs/1406.5823

Batson, D. C. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *Social Neuroscience of Empathy* (pp. 3–15). Cambridge, MA: MIT Press.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. doi:10.1111/spc3.12131

Baumeister, R. F., Vohs, K. D., Nathan DeWall, C., & Liqing Zhang. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, *11*(2), 167–203. doi:10.1177/1088868307301033

Bechara, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295. doi:10.1126/science.275.5304.1293

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372. doi:10.1016/j.geb.2004.06.010

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1-3), 7–15. doi:10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *The Journal of Neuroscience*, *18*(1), 428–437. doi:10.1.1.321.5317

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159–164. doi:10.1016/j.tics.2005.02.002

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–

981. doi:10.1287/opre.30.5.961

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, *33*(1), 1–27. doi:10.1287/opre.33.1.1

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. doi:10.1016/0010-0277(95)00676-P

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, *14*(4), 698–718. doi:10.1016/j.concog.2005.06.004

Blair, R. J. R. (2007a). Dysfunctions of medial and lateral orbitofrontal cortex in psychopathy. *Annals of the New York Academy of Sciences*, *1121*, 461–479. doi:10.1196/annals.1401.017

Blair, R. J. R. (2007b). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387–392. doi:10.1016/j.tics.2007.07.003

Blair, R. J. R. (2011). Moral judgment and psychopathy. *Emotion Review*, *3*(3), 296–298. doi:10.1177/1754073911406297

Blair, R. J. R. (2013). Psychopathy: cognitive and neural dysfunction. *Dialogues in Clinical Neuroscience*, *15*(2), 181–90.

Blair, R. J. R., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, *34*(2), 192–198. doi:10.1111/j.1469-8986.1997.tb02131.x

Blair, R. J. R., Sellars, C., Strickland, I., Clark, F., Williams, A., Smith, M., & Jones, L. (1996). Theory of Mind in the psychopath. *The Journal of Forensic Psychiatry*, *7*(1), 15–25. doi:10.1080/09585189608409914

Blair, R. J. R., White, S. F., Meffert, H., & Hwang, S. (2013). Emotional learning and the development of differential moralities: implications from research on psychopathy. *Annals of the New York Academy of Sciences*, *1299*, 36–41. doi:10.1111/nyas.12169/full

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, *11*(3), 118–125. doi:10.1016/j.tics.2006.12.006

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2015). *Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?* Retrieved from http://arxiv.org/abs/1510.03346

Braunack-Mayer, A. J., & Joy, A. (2001). What makes a problem an ethical problem? An empirical perspective on the nature of ethical problems in general practice. *Journal of Medical Ethics*, *27*(2), 98–103. doi:10.1136/jme.27.2.98

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. doi:10.1038/nrn756

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., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, *15*(5), 655–661. doi:10.1038/nn.3087

Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, *6*(10), 426–431. doi:10.1016/S1364-6613(02)01974-5

Camille, N. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, *304*(5674), 1167–1170. doi:10.1126/science.1094550

Carmona-Perera, M., Clark, L., Young, L. L., Pérez-García, M., & Verdejo-García, A. (2014). Impaired decoding of fear and disgust predicts utilitarian moral judgment in alcohol-dependent individuals. *Alcoholism: Clinical and Experimental Research*, *38*(1), 179–185. doi:10.1111/acer.12245

Carmona-Perera, M., Reyes del Paso, G. A., Pérez-García, M., & Verdejo-García, A. (2013). Heart rate correlates of utilitarian moral decision-making in alcoholism. *Drug and Alcohol Dependence*, *133*(2), 413–419. doi:10.1016/j.drugalcdep.2013.06.023

Carney, D. R., & Mason, M. F. (2010). Decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology*, *46*(4), 668–671. doi:10.1016/j.jesp.2010.02.003

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, *70*(3), 560–572. doi:10.1016/j.neuron.2011.02.056

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579–1601. doi:10.1111/j.1468-0262.2006.00719.x

Choe, S., & Min, K. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, *6*(7), 580–592.

Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience*, *5*(1), 59–67. doi:10.1093/scan/nsp051

Cleckley, H. (1976). *The mask of sanity.* (5th ed). St. Louis, MO: Mosby, Inc.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604–608. doi:10.1038/386604a0

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience*, *8*(9), 1255–1262. doi:10.1038/nn1514

Coricelli, G., & Rustichini, A. (2009). Reward-based emotions. In J.-C. Dreher & L. Tremblay (Eds.), *Handbook of Reward and Decision Making* (pp. 427–439). Elsevier. doi:10.1016/B978-0-12-374620-7.00020-0

Critchley, H. D., Corfield, D. R., Chandler, M. P., Mathias, C. J., & Dolan, R. J. (2000). Cerebral correlates of autonomic cardiovascular arousal: a functional neuroimaging investigation in humans. *The Journal of Physiology*, *523*(1), 259–270. doi:10.1111/j.1469-7793.2000.t01-1-00259.x

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(40), 17433–8. doi:10.1073/pnas.1009396107

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2003). The preparation and readiness for voluntary movement: A high-field event-related fMRI study of the Bereitschafts-BOLD response. *NeuroImage*, *20*(1), 404–412. doi:10.1016/S1053-8119(03)00291-X

Ćurčić-Blake, B., van der Meer, L., Pijnenborg, G. H. M., David, A. S., & Aleman, A. (2015).

Insight and psychosis: Functional and anatomical brain connectivity and self-reflection in Schizophrenia. *Human Brain Mapping*, n/a–n/a. doi:10.1002/hbm.22955

Cushman, F. A. (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. A. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292. doi:10.1177/1088868313495594

Cushman, F. A., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*, 2–7. doi:10.1037/a0025071

Cushman, F. A., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, 37–41. doi:10.1080/17470919.2011.614000

Cushman, F. A., & Young, L. L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*(6), 1052–1075. doi:10.1111/j.1551-6709.2010.01167.x

Cushman, F. A., Young, L. L., & Hauser, M. D. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089. doi:10.1111/j.1467-9280.2006.01834.x

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Avon Books.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*(2), 81–94. doi:10.1016/0166-4328(90)90144-4

Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science*, *5*(1), 1–23. doi:10.1146/annurev.lawsocsci.4.110707.172335

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*(4), 324–336. doi:10.1207/S15327957PSPR0704_05

Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*(1), 525–556. doi:10.1146/annurev.ps.41.020190.002521

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

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2288–2304. doi:10.1016/j.neubiorev.2012.08.006

de Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition & Emotion*, *21*(5), 1025–1042. doi:10.1080/02699930600980874

Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress, N. G. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology*, *115*(1), 131–144. doi:10.1037/0021-843X.115.1.131

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American*

Psychologist, *49*(8), 709–724. doi:10.1037/0003-066X.49.8.709

Epstude, K., & Roese, N. J. (2008). The Functional Theory of Counterfactual Thinking. *Personality and Social Psychology Review*, *12*(2), 168–192. doi:10.1177/1088868308316091

Ernst, M., & Paulus, M. P. (2005, October 15). Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biological Psychiatry*. doi:10.1016/j.biopsych.2005.06.004

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology*, *35*(12), 1731–1731. doi:10.1212/WNL.35.12.1731

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. doi:10.1146/annurev.psych.59.103006.093629

Foot, P. (1983). Moral realism and moral dilemma. *The Journal of Philosophy*, *80*(7), 379–398.

Fox, J. (2003). Eff ect displays in R for Generalised Linear Models. *Journal of Statistical Software*, *8*(15), 1–27. doi:10.2307/271037

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228. doi:10.1037/0022-3514.57.2.212

Frith, C. D., & Frith, U. (2005). Theory of mind. *Current Biology*, *15*(17), 644–645. doi:10.1016/j.cub.2005.08.041

Frith, U., & de Vignemont, F. (2005). Egocentrism, allocentrism, and Asperger syndrome. *Consciousness and Cognition*, *14*(4), 719–738. doi:10.1016/j.concog.2005.04.006

Gao, Y., & Tang, S. (2013). Psychopathic personality and utilitarian moral judgment in college students. *Journal of Criminal Justice*, *41*(5), 342–349. doi:10.1016/j.jcrimjus.2013.06.012

Giorgetta, C., Zeelenberg, M., Ferlazzo, F., & D'Olimpio, F. (2012). Cultural variation in the role of responsibility in regret and disappointment: The Italian case. *Journal of Economic Psychology*, *33*(4), 726–737. doi:10.1016/j.joep.2012.02.003

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

Glenn, A. L., Koleva, S., & Iyer, R. (2010). Moral identity in psychopathy. *Judgment and Decision Making*, *5*(7), 497–505.

Glenn, A. L., Raine, A., & Schug, R. a. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, *14*(1), 5–6. doi:10.1038/mp.2008.104

Glenn, A. L., Raine, A., Schug, R. A., Young, L. L., & Hauser, M. D. (2009). Increased DLPFC activity during moral decision-making in psychopathy. *Molecular Psychiatry*, *14*(10), 909–911. doi:10.1038/mp.2009.76

Gluth, S., Rieskamp, J., & Buchel, C. (2012). Deciding when to decide: Time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *Journal of Neuroscience*, *32*(31), 10686–10698. doi:10.1523/JNEUROSCI.0727-12.2012

Gluth, S., Rieskamp, J., & Büchel, C. (2013). Classic EEG motor potentials track the emergence of value-based decisions. *NeuroImage*, *79*, 394–403. doi:10.1016/j.neuroimage.2013.05.005

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 56–67. doi:10.1016/j.tics.2010.12.004

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. doi:10.1016/0013-4694(83)90135-9

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science*, *19*(12), 1260–1262. doi:10.1111/j.1467-9280.2008.02208.x

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development.* (pp. 35–79). Cambridge, MA: MIT Press.

Greene, J. D., Cushman, F. A., & Stewart, L. E. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371. doi:10.1016/j.cognition.2009.02.001

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work ?, *6*(12), 517–523.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. doi:10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. doi:10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108. doi:10.1126/science.1062872

Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry*, *52*(5), 449–466. doi:10.1080/00201740903302600

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences*, *9*(6), 290–295. doi:10.1016/j.tics.2005.04.012

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, *126*(1), 128–133. doi:10.1007/s002210050722

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. doi:10.1037//0033-295X.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). New York, NY: Oxford University Press.

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity* (pp. 181–218). Cambridge, MA: MIT Press.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–628.

Hare, R. D. (2003). *Manual for the Hare Psychopathy Checklist—Revised* (2nd ed). Toronto, Ontario, Canada: Multi-Health Systems.

Hauser, M. D., Cushman, F. A., Young, L. L., Kang-Xing Jin, R., & Mikhail, J. (2007). A

dissociation between moral judgments and justifications. *Mind & Language*, *22*(1), 1–21. doi:10.1111/j.1468-0017.2006.00297.x

Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *Journal of Neuroscience*, *35*(36), 12593–12605. doi:10.1523/JNEUROSCI.3402-14.2015

IBM Corp. (2010). IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Karpman, B. (1941). On the need of separating psychopathy into two distinct clinical types: the symptomatic and the idiopathic. *Journal of Criminal Psychopathology*, *3*, 112–137.

Ketelaar, T., & Tung Au, W. (2003). The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition & Emotion*, *17*(3), 429–453. doi:10.1080/02699930143000662

Khemiri, L., Guterstam, J., Franck, J., & Jayaram-Lindström, N. (2012). Alcohol dependence associated with increased utilitarian moral judgment: a case control study. *PloS One*, *7*(6), e39882. doi:10.1371/journal.pone.0039882

Koenigs, M., Kruepke, M., & Newman, J. P. (2010). Economic decision-making in psychopathy: a comparison with ventromedial prefrontal lesion patients. *Neuropsychologia*, *48*(7), 2198–2204. doi:10.1016/j.neuropsychologia.2010.04.012

Koenigs, M., Kruepke, M., Zeier, J. D., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, *7*(6), 708–714. doi:10.1093/scan/nsr048

Koenigs, M., Young, L. L., Adolphs, R., Tranel, D., Cushman, F. A., Hauser, M. D., & Damasio, A. R. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. doi:10.1038/nature05631

Kohlberg, L., Levine, C., & Hewer, A. (1983). *Moral stages: A current formulation and a response to critics*. Basel: Karger.

Kosslyn, S. M., Shin, L. M., Thompson, W. L., McNally, R. J., Rauch, S. L., Pitman, R. K., & Alpert, N. M. (1996). Neural effects of visualizing and perceiving aversive stimuli. *NeuroReport*, *7*(10), 1569–1576. doi:10.1097/00001756-199607080-00007

Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. 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

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, *33*(4), 323–333. doi:10.1016/j.evolhumbehav.2011.11.002

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). lmerTest: Tests in Linear Mixed Effects Models. Retrieved from https://cran.r-project.org/package=lmerTest

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system*

(IAPS): affective ratings of pictures and instruction manual. Technical Report A-7. University of Florida, Gainesville, FL.

Larsen, R. J., & Fredrickson, B. L. (1999). Measurement issues in emotion research. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 40–60). New York: Russell Sage.

Lee, D. (2004). Behavioral context and coherent oscillations in the Supplementary Motor Area. *Journal of Neuroscience*, *24*(18), 4453–4459. doi:10.1523/JNEUROSCI.0047-04.2004

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, *68*(1), 151–158. doi:10.1037/0022-3514.68.1.151

Levenston, G. K., Patrick, C. J., Bradley, M. M., & Lang, P. J. (2000). The psychopath as observer: Emotion and attention in picture processing. *Journal of Abnormal Psychology*, *109*(3), 373–385. doi:10.1037/0021-843X.109.3.373

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, *106*(3), 623–642.

Liljeholm, M., Dunne, S., & O'Doherty, J. P. (2014). Anterior insula activity reflects the effects of intentionality on the anticipation of aversive stimulation. *Journal of Neuroscience*, *34*(34), 11339–11348. doi:10.1523/JNEUROSCI.1126-14.2014

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(03), 121–143. doi:10.1017/S0140525X11000446

Loewenstein, G. F., & Lerner, J. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Science* (pp. 619–642). New York: Oxford University Press.

Loewenstein, G. F., Weber, E., Hsee, C., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. doi:10.1037/0033-2909.127.2.267

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824. doi:10.2307/2232669

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, *53*(2), 271–282. doi:10.2307/2297651

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making*, *27*(1), 57–65. doi:10.1002/bdm.1782

Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neurosciences*, *22*(7), 310–316. doi:10.1016/S0166-2236(98)01374-5

Makwana, A., Gron, G., Fehr, E., & Hare, T. A. (2015). A neural mechanism of strategic social choice under sanction-induced norm compliance. *eNeuro*, *2*(3), 1–8. doi:10.1523/ENEURO.0066-14.2015

Malle, B. F. (2015). Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*. doi:10.1007/s10676-015-9367-8

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial

populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *32*(3), 454–465. doi:10.1016/j.neubiorev.2007.08.003

Marsh, A. A., Finger, E. C., Schechter, J. C., Jurkowitz, I. T. N., Reid, M. E., & Blair, R. J. R. (2011). Adolescents with psychopathic traits report reductions in physiological responses to fear. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(8), 834–841. doi:10.1111/j.1469-7610.2010.02353.x

Masserman, J. H., Wechkin, S., & Terris, W. (1964). "Altruistic" behavior in rhesus monkeys. *The American Journal of Psychiatry*, *121*(6), 584–585.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*. doi:10.1509/jmkr.45.6.633

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*(3), 332–345. doi:10.1037/0096-3445.128.3.332

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology : Official Journal of the Society for Behavioral and Cognitive Neurology*, *18*(4), 193–197. doi:10.1097/01.wnn.0000191292.17964.bb

Mendez, M. F., Chen, A. K., Shapira, J. S., & Miller, B. L. (2005). Acquired sociopathy and frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders*, *20*(2-3), 99–104. doi:10.1159/000086474

Mikhail, J. (2002). Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect. *Georgetown Law and Economics Research Paper*, (762385).

Mikhail, J. (2011). Emotion, neuroscience, and law: A comment on Darwin and Greene. *Emotion Review*, *3*(3), 293–295. doi:10.1177/1754073911406150

Millar, J. C., Turri, J., & Friedman, O. (2014). For the greater goods? Ownership rights and utilitarian moral judgment. *Cognition*, *133*(1), 79–84. doi:10.1016/j.cognition.2014.05.018

Miller, R. M., & Cushman, F. A. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*(10), 707–718. doi:10.1111/spc3.12066

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, *14*(3), 573–587. doi:10.1037/a0035361

Moll, J., de Oliveira-Souza, R., & Eslinger, P. J. (2003). Morals and the human brain: A working model. *NeuroReport*, *14*(3), 299–305. doi:10.1097/00001756-200303030-00001

Moore, A. B., Clark, B., & Kane, M. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–558.

Moore, A. B., Stevens, J., & Conway, A. R. A. (2011). Individual differences in sensitivity to reward and punishment predict moral judgment. *Personality and Individual Differences*, *50*(5), 621–625. doi:10.1016/j.paid.2010.12.006

Moretto, G., Làdavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888–1899. doi:10.1162/jocn.2009.21367

Nadler, J., & McDonnell, M. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review*, *97*, 255–304.

Nakao, T., Ohira, H., & Northoff, G. (2012). Distinction between externally vs. internally guided decision-making: Operational differences, meta-analytical comparisons and their theoretical implications. *Frontiers in Neuroscience*, *6*(March), 31. doi:10.3389/fnins.2012.00031

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., … Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*(6), 1546–1554. doi:10.1212/WNL.51.6.1546

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*(2), 221–236. doi:10.1016/S0010-0277(02)00048-3

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*(3), 530–542. doi:10.1016/j.cognition.2005.07.005

Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research*, *49*(2), 100–110. doi:10.1111/j.1468-5884.2007.00337.x

Panasiti, M. S., Pavone, E. F., Mancini, A., Merla, A., Grisoni, L., & Aglioti, S. M. (2014). The motor cost of telling lies: Electrocortical signatures and personality foundations of spontaneous deception. *Social Neuroscience*, (July), 1–17. doi:10.1080/17470919.2014.934394

Patil, I. (2015). Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion. *Journal of Cognitive Psychology*, *27*(3), 349–366. doi:10.1080/20445911.2015.1004334

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

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology*, *31*(4), 319–330. doi:10.1111/j.1469-8986.1994.tb02440.x

Paulus, M. P. (2005). Neurobiology of decision-making: Quo vadis? *Cognitive Brain Research*, *23*(1), 2–10. doi:10.1016/j.cogbrainres.2005.01.001

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163–177. doi:10.1111/j.1551-6709.2011.01210.x

Praamstra, P., Stegeman, D. F., Horstink, M. W. I. M., & Cools, A. R. (1996). Dipole source analysis suggests selective modulation of the supplementary motor area contribution to the readiness potential. *Electroencephalography and Clinical Neurophysiology*, *98*(6), 468–477. doi:10.1016/0013-4694(96)95643-6

Preston, S. D., & de Waal, F. B. M. (2001). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(01), 1–71. doi:10.1017/S0140525X02000018

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations*, *9*(1), 29–43. doi:10.1080/13869790500492466

Psychology Software Tools Inc. (2012). E-Prime 2.0. Retrieved from http://www.pstnet.com

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current*

Opinion in Neurobiology, *20*(2), 262–270. doi:10.1016/j.conb.2010.03.001

Reiman, E. M. (1997). The application of positron emission tomography to the study of normal and pathologic emotions. *The Journal of Clinical Psychiatry*, *58 Suppl 1*, 4–12.

Reiman, E. M., Lane, R. D., Ahern, G. L., Schwartz, G. E., Davidson, R. J., Friston, K. J., … Chen, K. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *American Journal of Psychiatry*, *154*(7), 918–925. doi:10.1176/ajp.154.7.918

Rice, G. E., & Gainer, P. (1962). "Altruism" in the albino rat. *Journal of Comparative and Physiological Psychology*, *55*(1), 123–125. doi:10.1037/h0042276

Rigoni, D., Kühn, S., Sartori, G., & Brass, M. (2011). Inducing disbelief in free will alters brain correlates of preconscious motor preparation: the brain minds whether we believe in free will or not. *Psychological Science*, *22*(5), 613–8. doi:10.1177/0956797611405680

Rösler, F., & Heil, M. (1991). Toward a Functional Categorization of Slow Waves: Taking into Account Past and Future Events. *Psychophysiology*, *28*(3), 344–358. doi:10.1111/j.1469-8986.1991.tb02205.x

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, *76*(4), 574–586. doi:10.1037/0022-3514.76.4.574

Rushworth, M. F. S., Walton, M. E., Kennerley, S., & Bannerman, D. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, *8*(9), 410–417. doi:10.1016/j.tics.2004.07.009

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in the newborn. *Developmental Psychology*, *12*(2), 175–176. doi:10.1037/0012-1649.12.2.175

Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., & Nichols, T. E. (2009). Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. *NeuroImage*, *45*(3), 810–823. doi:10.1016/j.neuroimage.2008.12.039

Sanfey, A. G., Rilling, J. K., Aronson, J. a, Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science (New York, N.Y.)*, *300*(5626), 1755–1758. doi:10.1126/science.1082976

Sarlo, M., Lotto, L., Manfrinati, A., Rumiati, R., Gallicchio, G., & Palomba, D. (2012). Temporal dynamics of cognitive-emotional interplay in moral decision-making. *Journal of Cognitive Neuroscience*, *24*(4), 1018–1029. doi:10.1162/jocn_a_00146

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology & Behavior*, *130C*, 127–134. doi:10.1016/j.physbeh.2014.04.002

Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, *18*(5), 803–817. doi:10.1162/jocn.2006.18.5.803

Schleim, S., Spranger, T. M., Erk, S., & Walter, H. (2011). From moral to legal judgment: The influence of normative context in lawyers and other academics. *Social Cognitive and Affective Neuroscience*, *6*(1), 48–57. doi:10.1093/scan/nsq010

Seara-Cardoso, A., Dolberg, H., Neumann, C., Roiser, J. P., & Viding, E. (2013). Empathy, morality and psychopathic traits in women. *Personality and Individual Differences*, *55*(3),

328–333. doi:10.1016/j.paid.2013.03.011

Seara-Cardoso, A., Neumann, C., Roiser, J., McCrory, E., & Viding, E. (2012). Investigating associations between empathy, morality and psychopathic personality traits in the general population. *Personality and Individual Differences*, *52*(1), 67–71. doi:10.1016/j.paid.2011.08.029

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience*, *34*(13), 4741–4749. doi:10.1523/JNEUROSCI.3390-13.2014

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *117*(11), 2341–2356. doi:10.1016/j.clinph.2006.04.025

Shultz, T. R., & Wells, D. (1985). Judging the intentionality of action-outcomes. *Developmental Psychology*, *21*(1), 83–89. doi:10.1037/0012-1649.21.1.83

Simpson, J., Hillman, R., Crawford, T., & Overton, P. G. (2010). Self-esteem and self-disgust both mediate the relationship between dysfunctional cognitions and depressive symptoms. *Motivation and Emotion*, *34*(4), 399–406. doi:10.1007/s11031-010-9189-2

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*, 81–96. doi:10.1111/j.1749-6632.2009.04418.x

Sinnott-Armstrong, W. (1987). Moral realisms and moral dilemmas. *The Journal of Philosophy*, *84*(5), 263–276.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., & Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience*, *7*(1), 80–84. doi:10.1038/nn1160

Skeem, J., Johansson, P., Andershed, H., Kerr, M., & Louden, J. E. (2007). Two subtypes of psychopathic violent offenders that parallel primary and secondary variants. *Journal of Abnormal Psychology*, *116*(2), 395–409. doi:10.1037/0021-843X.116.2.395

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. doi:10.1.1.121.5355

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42. doi:10.1006/cogp.1997.0658

Sood, S., & Forehand, M. (2005). On self-referencing differences in judgment and choice. *Organizational Behavior and Human Decision Processes*, *98*(2), 144–154. doi:10.1016/j.obhdp.2005.05.005

Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the neural correlates of schemas: Ventromedial prefrontal cortex is necessary for normal schematic influence on memory. *Journal of Neuroscience*, *35*(47), 15746–15751. doi:10.1523/JNEUROSCI.2767-15.2015

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: Chicago University Press.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421–440). New York: Cambridge University Press.

Sui, J., Enock, F., Ralph, J., & Humphreys, G. W. (2015). Dissociating hyper and hypoself biases to a core self-representation. *Cortex*, *70*, 202–212. doi:10.1016/j.cortex.2015.04.024

Szekely, R. D., & Miu, A. C. (2015). Incidental emotions in moral dilemmas: The influence of emotion regulation. *Cognition and Emotion*, *29*(1), 64–75. doi:10.1080/02699931.2014.895300

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, *70*(6), 1256–1269. doi:10.1037/0022-3514.70.6.1256

Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013). High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in Human Neuroscience*, *7*, 1–6. doi:10.3389/fnhum.2013.00229

Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., … Wicker, B. (2012). Disrupting the right prefrontal cortex alters moral judgement. *Social Cognitive and Affective Neuroscience*, *7*(3), 282–288. doi:10.1093/scan/nsr008

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between Judgment and Choice of Action in Moral Dilemmas. *Frontiers in Psychology*, *4*(May), 250. doi:10.3389/fpsyg.2013.00250

Terbek, S., Guy, K., Sarah, M., Julian, S., Neil, L., Miles, H., & Cowen, P. J. (2013). Beta adrenergic blockade reduces utilitarian judgement. *Biological Psychology*, *92*(2), 323–328. doi:10.1016/j.biopsycho.2012.09.005

Thomson, J. (1985). The Trolley Problem. *Yale Law Journal*, *94*, 1395–1415.

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

Turiel, E. (2002). *The Culture of Morality*. Cambridge, UK: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. doi:10.1.1.304.6549

Ugazio, G., Lamm, C., & Singer, T. (2012). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion*, *12*(3), 579–590. doi:10.1037/a0024611

Valdesolo, P., & DeSteno, D. (2006). Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science*, *17*(6), 476–477. doi:10.1111/j.1467-9280.2006.01731.x

Vanman, E. J., Mejia, V. Y., Dawson, M. E., Schell, A. M., & Raine, A. (2003). Modification of the startle reflex in a community sample: do one or two dimensions of psychopathy underlie emotional processing? *Personality and Individual Differences*, *35*(8), 2007–2021. doi:10.1016/S0191-8869(03)00052-7

Vera-Estay, E., Dooley, J. J., & Beauchamp, M. H. (2014). Cognitive underpinnings of moral reasoning in adolescence: The contribution of executive functions. *Journal of Moral Education*, (February 2015), 1–17. doi:10.1080/03057240.2014.986077

Verona, E., Patrick, C. J., Curtin, J. J., Bradley, M. M., & Lang, P. J. (2004). Psychopathy and physiological response to emotionally evocative sounds. *Journal of Abnormal Psychology*, *113*(1), 99–108. doi:10.1037/0021-843X.113.1.99

Wager, T. D., Lindquist, M. A., Nichols, T. E., Kober, H., & Van Snellenberg, J. X. (2009). Evaluating the consistency and specificity of neuroimaging data using meta-analysis.

NeuroImage, *45*(1), S210–S221. doi:10.1016/j.neuroimage.2008.10.061

Wagner, U., Handke, L., Dörfel, D., & Walter, H. (2012). An experimental decision-making paradigm to distinguish guilt and regret and their self-regulating function via loss averse choice behavior. *Frontiers in Psychology*, *3*(1995), 431. doi:10.3389/fpsyg.2012.00431

Wagner, U., N'Diaye, K., Ethofer, T., & Vuilleumier, P. (2011). Guilt-specific processing in the prefrontal cortex. *Cerebral Cortex*, *21*(11), 2461–2470. doi:10.1093/cercor/bhr016

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780–784.

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition*, *126*(3), 364–372. doi:10.1016/j.cognition.2012.11.002

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting. Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131–134. doi:10.1111/j.0963-7214.2005.00355.x

Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences*, *106*(40), 17199–17204. doi:10.1073/pnas.0901077106

Young, L. L., Bechara, A., Tranel, D., Damasio, H., Hauser, M. D., & Damasio, A. R. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*(6), 845–851. doi:10.1016/j.neuron.2010.03.003

Young, L. L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, *7*(1), 1–10. doi:10.1080/17470919.2011.569146

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

Young, L. L., & Saxe, R. (2009a). An fMRI Investigation of Spontaneous Mental State Inference for Moral Judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405. doi:10.1162/jocn.2009.21137

Young, L. L., & Saxe, R. (2009b). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*(10), 2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020

Zeelenberg, M. (2008). On emotion specificity in decision making: Why feeling is for doing. *Judgment and Decision Making*, *3*(1), 18–27.

Zeelenberg, M., & Pieters, R. (1999). Comparing service delivery to what might have been: Behavioral responses to regret and disappointment. *Journal of Service Research*, *2*(1), 86–97. doi:10.1177/109467059921007

Zeelenberg, M., van der Pligt, J., & de Vries, N. K. (2000). Attributions of responsibility and affective reactions to decision outcomes. *Acta Psychologica*, *104*, 303–315. doi:10.1016/S0001-6918(00)00034-2

Zeelenberg, M., van Dijk, W. W., & Manstead, A. (2000). Regret and responsibility resolved? Evaluating Ordóñez and Connolly's (2000) conclusions. *Organizational Behavior and Human Decision Processes*, *81*(1), 143–154. doi:10.1006/obhd.1999.2865