Moral Reasoning about Human Welfare in Adolescents and Adults:
Judging Conflicts Involving Sacrificing and Saving Lives

Manuscript in press in *Monographs of the Society for Research in Child Development*.

Audun Dahl
University of California, Santa Cruz

Matthew Gingo
Wheaton College

Kevin Uttich
University of California, Berkeley

Elliot Turiel
University of California, Berkeley

**Author note**

Correspondence concerning this article should be addressed to Audun Dahl, Department of Psychology, University of California, Santa Cruz, CA 95064. E-mail: dahl@ucsc.edu
Table of Contents

Abstract ........................................................................................................................................... 1
Chapter 1: Introduction ..................................................................................................................... 2
Chapter 2: Study 1: Investigating both Evaluations and Reasoning about Systematically Varied Trolley Car Situations ........................................................................................................ 36
Chapter 3: Study 2: Varying the Involvement of the Potential Victims ........................................... 76
Chapter 4: General Discussion ........................................................................................................... 91
References ....................................................................................................................................... 124
Acknowledgements .......................................................................................................................... 141
Contributors ................................................................................................................................. 142
Tables ............................................................................................................................................. 143
Abstract

The value of life is a significant moral value for most people. Yet, past research has devoted little attention to the development of moral reasoning about the value of life. The present studies investigated how adolescents and adults reason about the value of life in the context of so-called *trolley car situations*. These situations, adopted from philosophy, involve the option of sacrificing the life of one person to save five others. Based on past developmental research, we expected that individuals would reason about distinct and sometimes conflicting considerations regarding the value of life. This approach contrasted with past research on adults’ responses to trolley car situations, which has been taken to show that most moral evaluations are based not on reasoning but on affective, automatic reactions. In Study 1, 288 adolescents and adults were interviewed about trolley car situations designed to examine considerations like the value of human life and the relationship of those at risk with the actors. In Study 2, 144 college students were interviewed to further examine the roles of those involved. Participants’ justifications referred not only to the number of lives saved, but also to other considerations, such as intrinsic rights and personal responsibility for events. Moreover, responses indicated frequent conflicts about standard trolley car situations, counter to the argument that people’s evaluations are automatic or based solely on a counting of lives saved. The present findings indicate that adolescents and adults reason about, and seek to coordinate, distinct moral considerations regarding the value of life.

*Keywords:* moral reasoning; moral evaluations; moral development
Chapter 1: Introduction

The value of life is a significant moral value for most people (Dworkin, 1993; Kohlberg, 1963; Turiel & Killen, 2010). Moral concerns with life and death are evident by adolescence (Balk & Corr, 2009). The taking and saving of lives, though not necessarily directly confronted by most adolescents and adults, are issues encountered in literature and in media coverage of war, capital punishment, and violent crime (Anderson, 1999; Balk & Corr, 2009; Posada & Wainryb, 2008). Without doubt, public discussions and legal matters pertaining to the preservation and taking of lives are within the awareness of most individuals. These issues can present complex problems because they often involve competing considerations. One question often in the public light is whether to administer capital punishment, which can involve conflicting considerations regarding the value of life, deterrence of future crimes, and retributive justice. An issue first raised in philosophical discourse and later in psychological research is whether it is permissible to sacrifice one life to save others. The philosophical literature suggests that evaluations about taking and saving lives may be guided not only by the intrinsic rights of human life, but also by the weighting of one life against other lives, the roles and responsibilities of the potential victims, and the entitlements of outsiders to sacrifice or save lives (see e.g. Edmonds, 2013; Korsgaard, 1996; Wiggins, 2006; Williams, 1973).

This Monograph deals with the development of moral reasoning about taking and saving lives. Many adolescents and adults think that taking a life is sometimes acceptable but that killing is generally wrong, suggesting that they are sensitive to competing considerations regarding the value of life (Dworkin, 1993; Greene, 2014; Turiel, Hildebrandt, & Wainryb, 1991). Still, past research has not systematically investigated how adolescents and adults reason
about the competing considerations involved when the life of one person is pitted against the life of others.

The present research is based on the proposition that reasoning is central to how children, adolescents, and adults approach moral issues (Killen & Smetana, 2015; Kohlberg, 1963; Turiel, 2015; Turiel & Dahl, in press; Turiel & Killen, 2010). By this, we mean that evaluations, decisions, and actions about moral issues are grounded in considerations about harm, welfare, rights, fairness, and justice. By preschool age, children evaluate acts of harm as wrong and base these evaluations on reasoning about the need to avoid hurting people (Davidson, Turiel, & Black, 1983; Nucci & Turiel, 2009; Nucci & Weber, 1995; Smetana, 2013; Turiel, 1983, 2015). Within our framework, reasoning operates in concert with, not in opposition to, emotions: People are emotional about issues they reason about and reasoning can give rise to emotional reactions (Lazarus, 2001; Piaget, 1981; Turiel & Dahl, in press; Turiel & Killen, 2010).

In multifaceted situations, for instance pitting the rights of one person against the welfare of another, individuals coordinate and prioritize competing moral and non-moral considerations (Turiel, 2008a; Turiel et al., 1991). Over the course of development, responses to multifaceted situations reveal both continuities and changes; continuities because fundamental moral concerns with welfare, fairness, and rights develop quite early and are maintained at later ages, changes occur because experiences with multifaceted situations lead individuals to coordinate those fundamental moral concerns in new ways (Kohlberg, 1963; Nucci & Turiel, 2009; Turiel & Nucci, 2017).

The proposition that reasoning is central to most aspects of morality differs from recent dual-process and intuitionist views (Greene, 2013, 2014, Haidt, 2001, 2007). According to the latter theories, most moral evaluations are based on automatic emotional reactions rather than
reasoning. As stated by Greene, “We rely on our automatic settings most of the time” (Greene, 2014, p. 696; Paxton & Greene, 2010). The view that automatic, unconscious, and affective reactions pervade morality is partly based on the claim that people are often unable to justify their moral evaluations, a phenomenon called “moral dumbfounding” (Cushman, Young, & Hauser, 2006; Haidt, Bjorlund, & Murphy, 2000; Hauser, Cushman, Young, Kang-Xing Jin, & Mikhail, 2007). It is also claimed that even when people do try to reason about moral issues, their reasoning is often overridden by strong emotional reactions or relegated to supplying post hoc rationalizations for automatic evaluations already made (Greene, 2014). Finally, some dual-process theorists argue that moral reasoning, when it does occur, mainly amounts to the application of the utilitarian principle of maximizing total welfare, for instance by sacrificing one life to save five (Greene, 2014).

The dual-process and intuitionist views are largely based on research with adults, but have implications for theories of moral development based on findings that children form complex configurations of moral thinking (Turiel, 2105). If adults do not typically reason about moral issues, adolescents must either lose their propensity for moral reasoning or actually had little such propensity to begin with, challenging past developmental research. Moreover, if many moral evaluations are based on emotional reactions that operate independently from reasoning, developmental research would be needed to explain how these independent emotional reactions develop. This Monograph aimed to address discrepancies between reasoning-based perspectives on moral development and dual-process perspectives on adult moral psychology.

In the present research, two lines of inquiry – one regarding the development of orientations toward taking and saving lives, the other regarding the role of reasoning in morality – converged in an investigation of how adolescents and adults respond to a set of philosophical
problems known as *trolley car situations* (Edmonds, 2013; Foot, 1967; Thomson, 1976). These situations center on a protagonist’s decision about whether to sacrifice the life of one person in order to save the lives of several others. Not only are these situations thereby well suited to investigate how adolescents and adults approach complex moral issues involving taking and saving lives; these trolley car situations have also been used in research taken to support the dual process and intuitionist approaches to morality (Greene, 2013, 2014; Haidt, 2007; Kahane, 2015).

In the standard pair of trolley car situations – labeled *switch* and *footbridge* – a runaway trolley car is headed toward five railway workmen on a train track who will be killed if a bystander does not intervene. The only way to save the lives of the five others is to sacrifice the life of one person. In a so-called *switch* variant, the only way of saving the five persons is to flip a switch, which redirects the train onto a sidetrack where it will kill one workman. In this situation, most people say it is permissible to intervene, sacrificing the life of one person to save the lives of five others (Greene et al., 2009; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Hauser et al., 2007). In a *footbridge* variant, the protagonist has the option of pushing a stranger off a footbridge onto the tracks, thereby killing the person pushed but stopping the train before it hits the five persons on the track. In contrast to the switch situation, most people say it is not permissible to intervene in the footbridge situation.

Proponents of intuitionist or dual process views have taken adults’ evaluations of trolley car situations to support the claim that many moral decisions are based on automatic emotional reactions. According to this interpretation, affective and non-reasoned evaluations are most common in response to the footbridge situation (Greene, 2014; Greene et al., 2009). When they imagine pushing the person off the footbridge, people are said to have an automatic, negative
affective reaction to the idea of pushing, which in turn leads them to say it is wrong to intervene in the footbridge situation. It is also proposed that people do reason about the switch situation, but only insofar as they apply utilitarian principles by comparing the number of lives lost and saved. Since flipping a switch is said to elicit no affective reaction, individuals are free to follow their utilitarian reasoning, according to the dual-process view.

Our contrasting view that reasoning is involved in most moral evaluations – including non-utilitarian ones – leads to a different explanation of how individuals respond to trolley car situations. On this account, responses to the standard trolley car situations do not involve a competition between automatic emotional reactions and reasoning; rather, people’s evaluations reflect reasoning about moral concerns that are pitted against each other by the design of the trolley situations. Based on this account, the present research investigated the following hypotheses about how adolescents and adults would respond to trolley car situations: (1) First, we proposed that individuals would articulate several moral considerations about the trolley car situations, not just utilitarian reasons about maximizing the number of lives saved. (2) Next, we proposed that individuals would be conflicted about both the switch and footbridge variants, and that neither variant would typically be judged effortlessly but rather elicit a multitude of conflicting considerations. (3) Lastly, we proposed that the evaluations and reasons provided by participants would be sensitive to the features of the situations under evaluation. That is, we expected people’s evaluations and reasons or justifications to change in consistent ways in response to manipulations of morally relevant features of the situations.

It was an open question whether developmental continuities or developmental changes would be most prominent in comparisons of adolescent and adult responses to trolley situations. The idea of development, of course, points in the direction of change, but we know from a good
deal of research that continuities are also important and provide the basis for developmental differences. For instance, there is substantial continuity in conceptions of interpersonal harm from early childhood to adulthood: Children, adolescents, and adults say it is wrong to harm others in most situations (Turiel, 2015). Illustrating such continuity, one study found that virtually all 8- to 16-year-olds said it was wrong for a protagonist to hit another child for no reason, or even to hit a child with mental illness who had hit the protagonist first (Nucci & Turiel, 2009). At the same time, this study showed developmental discontinuity in response to certain contexts involving conflicting concerns: When asked whether a protagonist had the right to hit back after being hit by a non-disabled child, 14-year-olds were more likely than both 8-year-olds and 16-year-olds to say that the protagonist had the right to hit back. To explain these developmental changes, it is necessary to consider the underlying developmental continuities (e.g. virtually everyone thinks it is generally wrong to harm others) as well as the contexts in which developmental changes are found (e.g. contexts involving certain conflicts between evaluative concepts).

The Monograph reports the findings from two studies. In Study 1, we investigated evaluations, reasoning, and conflicts about trolley situations among adolescents and adults. Study 1 addressed all three hypotheses listed above. Study 2 was a follow-up study designed to further investigate the third hypothesis about responsiveness of evaluations and reasons to situational variations. Both studies differed from prior research on trolley car situations by using semi-structured interviews, entailing a series of standard questions aimed at uncovering salient features of participants’ thinking (Damon, 1977; Kohlberg, 1963; Piaget, 1929; Turiel, 1983). The present research builds on research on moral development, moral philosophical work
Moral reasoning relating to the trolley situations, and recent debates regarding dual process and intuitionist explanations of moral decisions.

The Development of Moral Evaluations: Continuity and Change

Developing evaluations, reasoning, and emotions in the moral domain. We view reasoning, in conjunction with emotions, as central to the formation of evaluations about moral and other normative issues. Based on philosophical work (e.g. Dworkin, 1977; Nussbaum, 2000; Rawls, 1971; Sen, 1999, 2009) and a large body of empirical research, we define moral issues as those pertaining to others’ welfare, rights, fairness, and justice (Smetana, 2013; Turiel, 1983, 1998, 2002). Moral issues thus defined differ from social conventional issues (pertaining to social organization), prudential issues (pertaining to the agent’s own welfare), and personal issues (issues deemed under personal jurisdiction).

A great deal of research on the development of moral evaluations has consistently shown that relatively young children, as well as older children and adolescents, evaluate acts of harm as wrong and base these evaluations on reasons related to the need to avoid hurting people (Davidson, Turiel, & Black, 1983; Nucci & Turiel, 2009; Nucci & Weber, 1995; Smetana, 2013; Turiel, 1983, 2015; Turiel & Nucci, 2017). Moreover, those studies found that children make consistent and cohesive evaluations about actions in the moral domain: they judge moral transgressions, such as arbitrarily hitting someone or taking the property of others, as wrong and regard such acts as not contingent on existing rules, or dictates of authority, or on existing common practices. The research also shows that such non-contingent and generalizable evaluations about moral issues are connected to reasoning regarding welfare, or avoidance of harm and pain to people, as well as considerations of fairness, justice, and rights (for reviews, see Killen & Smetana, 2015; Smetana, Jambon, & Ball, 2014; Turiel, 2002, 2015).
A number of studies have also shown that children form concepts of equality and fairness regarding the distribution of goods (Elenbaas, Rizzo, Cooley, & Killen, 2016; Olson & Spelke, 2008), and that in late childhood considerations of equity are increasingly brought into such decisions (Damon, 1977, 1980; Piaget, 1965). Furthermore, it has been shown that children grasp concepts of fairness in social inclusion and exclusion in groups (Killen, Pisacane, Lee-Kim, & Ardila-Rey, 2001) and the importance of upholding rights like freedoms of speech and religion and rights to nurturance (Helwig, 1995; Ruck, Abramovitch, & Keating, 1998).

We have proposed that moral and other evaluations develop through everyday interactions in the early years of life (Dahl & Freda, 2017; Turiel, 1983; Turiel & Dahl, in press). These experiences include direct experiences with harming others and being harmed, as well as emotional, physical, and verbal signals from caregivers (Dahl, 2016a, 2016b; Dahl & Campos, 2013; Dahl, Sherlock, Campos, & Theunissen, 2014; Nucci & Turiel, 1978; Smetana, 1984). Although some researchers have proposed that infants show preferences for helpful over hindering characters (Bloom, 2013; Hamlin, 2013; Hamlin, Wynn, & Bloom, 2007; for a discussion, see Dahl, 2014), it is not until the third and fourth years that children express categorical negative evaluations of others’ actions, for instance through protests or evaluations of hypothetical events (Rakoczy, Warneken, & Tomasello, 2008; Schmidt, Rakoczy, & Tomasello, 2012; Schmidt & Tomasello, 2012; Smetana et al., 2012; Smetana & Braeges, 1990).

By preschool age, children distinguish evaluations in the moral domain from evaluations in the domains of social convention and personal jurisdiction (Dahl & Kim, 2014; Killen & Smetana, 2015; Nucci & Weber, 1995; Smetana & Braeges, 1990; Turiel, 1983, 2015). When explaining why it is wrong to violate social conventions, such as dress codes, children reference authority commands, existing rules, and social coordination. Children also view these
Moral reasoning

conventions as alterable by authorities and non-generalizable, unlike moral rules. In contrast, preschoolers and older children view many issues, such as choice of clothing or play activity, as falling under personal jurisdiction and as not legitimately regulated by authorities or rules.

Moral evaluations are closely tied with the elicitation of emotions, starting in childhood (Turiel & Dahl, in press; Turiel & Killen, 2010). A number of theorists have maintained that emotions are evoked when individuals perceive events in the world as relevant to their goals, desires, and other concerns (Barrett & Campos, 1987; Lazarus, 1991; Moors & Scherer, 2013; Nussbaum, 2001). Insofar as people are concerned with moral issues – and want others to do what is right – people have emotional reactions when they perceive actions that violate principles for how to treat others (e.g. Bonanno & Keltner, 2004; Cushman, Gray, Gaffey, & Mendes, 2012; Dahl et al., 2014; Hoffman, 2000). On this view, moral reasoning and evaluations are integral to many emotions. (As we discuss later in this introduction, our view that emotions operate in concert with reasoning, not in opposition to it, contrasts with intuitionist and dual process views of the relation between emotion and reasoning.)

Developing orientations toward multifaceted situations. The research on young children’s reasoning about moral issues has largely used what have been referred to as straightforward situations, involving a unitary moral feature. For instance, children have been presented with a person hitting or stealing from another without provocation and for no other apparent reason than self-interest. As noted, there is continuity in responses to straightforward moral situations, in the sense that preschoolers, older children, adolescents, and adults evaluate and reason about these situations in similar ways.

In contrast, multifaceted situations involve multiple moral and non-moral features that are sometimes in conflict (Turiel, 1989, 2008a). For example, research (Helwig, 1995) has examined
evaluations about situations posing conflicts between rights (e.g., to free speech) and potential harm (e.g., giving a speech to incite violence), as well as conflicts between honesty and preventing harm (e.g. using deception in order to prevent harm to someone, Freeman, Rathore, Weinfurt, Schulman, & Sulmasy, 1999; Perkins & Turiel, 2007). These studies show that people’s decisions often involve a weighing and balancing, or coordinating, different moral considerations, as well as moral and non-moral considerations (Turiel, 2015).

The reasons people provide for evaluations about multifaceted situations can refer to a multiplicity of considerations, even within the moral domain, such as rights (e.g. free speech and personal freedom), psychological or physical welfare, and principles of fairness (e.g. considerations of merit, need, and equality) (Damon, 1975; Helwig, 2006; Helwig, Hildebrandt, & Turiel, 1995; Killen et al., 2001; Smetana, 2013; Wainryb, Brehl, & Matwin, 2005). In multifaceted situations, reasoning also involves coordinating different considerations (Turiel, 2002; Turiel et al., 1991).

Developmental differences appear more commonly in responses to multifaceted situations than in responses to straightforward situations. Above, we mentioned the example of the study which yielded developmental differences in evaluations about hitting in self-defense, but not in response to unprovoked hitting (Nucci & Turiel, 2009). Other examples of developmental differences in multifaceted situations include evaluations about resource distribution (e.g. pitting merit against equality, Damon, 1975; Elenbaas et al., 2016; Rizzo & Killen, 2016) and rights (Helwig, 1995). Our interpretation is that changes in evaluations about multifaceted situations reflect developmental changes in how children coordinate competing moral and non-moral considerations. Of course, this is not to say that all multifaceted situations will elicit developmental differences, only that differences appear to be more common in
Moral reasoning

situations requiring coordination between considerations (multifaceted) than in situations involving a single moral or non-moral consideration (straightforward).

**Moral concerns with the value of life from adolescence to adulthood:**

**Straightforward and multifaceted situations.** Evaluations and reasons about promoting welfare and preventing harm extend to the preservation of life: A strongly held evaluation across cultures is that life is valuable and sacred (Turiel, 2002; Turiel & Killen, 2010). For example, Dworkin (1993) maintained that people everywhere are concerned with preserving lives, act to save lives when they can, make evaluations about the loss of life, and experience intense emotions like grief at the loss of a loved one (see also Nussbaum, 2001). These moral concerns with the value of life are evident by adolescence, if not before. Past research has indicated that adolescents judge killing as wrong (Turiel et al., 1991) and are concerned with and affected by issues of life and death (Balk & Corr, 2009; Kohlberg, 1971).

Despite the near-universal agreement that human life is valuable, people do not always think it is wrong to take lives or obligatory to save lives. In fact, several phenomena of major individual and societal importance involve the deliberate taking of lives, such as self-defense, capital punishment, and war. As noted above, that people sometimes judge it acceptable to take a life does not mean that they do not value life or hold it sacred (Dworkin, 1993). Although many situations can be defined as entailing straightforward considerations regarding life (e.g., murdering a security guard to rob a bank), other situations involve conflicts (e.g., maiming or killing a bank robber to prevent the robber from murdering the security guard). These types of conflicts put a strongly held value (the value of life) in conflict with itself (taking a life to save a life).
The philosophical literature provides some suggestions as to why conflicts involving saving and taking lives can be challenging: There are a number of proposed principles that pertain to the value of life, and these principles sometimes lead to different conclusions. For instance, one of Kant’s (1785) formulations of the categorical imperative was that you should never use persons only as a means, and always also as an end in themselves. This would imply that you cannot use one person’s life to save the lives of others. In contrast, the utilitarian principle of acting in a way that maximizes the total welfare does prescribe the sacrifice of one person to save others under some circumstances (for a discussion, see Wiggins, 2006). In addition to these classic examples, a number of other moral considerations have been brought to bear on conflicts involving taking and saving lives, such as the distinction between negative and positive duties, the intrinsic value of life, and notions of personal integrity (Dworkin, 1993; Edmonds, 2013; Foot, 1967; Thomson, 2008; Williams, 1973).

There has been little research on the development of moral reasoning about life and death situations, in which the life or welfare of one person is pitted against the lives of others. It is important to do so, given that the value of life is both a fundamental moral concern and one that most people think it is sometimes permissible to violate. A further developmental challenge regarding the value of life is that many adolescents and adults develop views about taking and saving lives without themselves having direct experiences with such actions. One of few developmental investigations of reasoning about the value of life is that of Kohlberg and colleagues (Kohlberg, 1963; Colby & Kohlberg, 1987a, 1987b). Their work examined evaluations about life through situations that presented conflicts between life and property rights, as well as between life and adhering to the law. His analyses showed that by adolescence there are concerns with the sanctity of life, life as associated with respect for persons, and human
dignity. However, Kohlberg (1971) also argued that many adolescents, and even adults (i.e. those who did not reach his proposed most advanced stages of moral development), viewed the value of life as contingent on affection from friends and family or belongingness to a social group.

Kohlberg’s characterizations of the development of moral evaluations have subsequently been criticized (e.g. Turiel, 2008b). In particular, his claims that children at first confuse moral issues of welfare and rights with social conventions have been challenged by the research discussed above showing that preschoolers and older children do distinguish moral from conventional issues (Nucci & Weber, 1995; Smetana & Braeges, 1990; Turiel, 1983). Furthermore, one of the limitations of Kohlberg’s approach was that his analyses focused on the conclusions individuals reached about a situation, and not on any additional considerations individuals were seeking to coordinate.

In the present approach to studying how adolescents and adults reason about conflictful life and death situations, different from that of Kohlberg, an overarching goal was to formulate procedures that would capture the multiplicity of participants’ concerns. That is, we manipulated features of the situations so as to allow participants to express conflicting considerations, for instance both utilitarian and non-utilitarian considerations. Moreover, the interview and coding protocols were designed to elicit and record the extent to which participants were conflicted about their evaluations.

While our research aimed at capturing the multiple considerations individuals bring to bear on evaluations about saving and taking lives, it was also aimed at revealing developmental continuities and differences. However, past research did not yield strong reasons for hypotheses about whether continuities or differences would be more prominent. On the one hand, past research on moral development has found a great deal of continuity in reasoning about moral
issues. On the other hand, past research on certain conflictful situations has documented developmental changes in how individuals coordinate competing considerations to come to decisions.

The research reported here focused on the period from adolescence to adulthood. Past research has indicated that issues of life and death are salient to most adolescents (Balk & Corr, 2009). Moreover, children’s conceptions about death appear to change during childhood and into adolescence, although even young children have some understanding of death (Rosengren et al., 2014; Slaughter, 2005; Talwar, Harris, & Schleifer, 2011). We did not include young children because of ethical concerns about interviewing them about acts of taking lives before knowing how adolescents would respond to these interviews.

The trolley car situations are well suited for investigations of evaluations and reasoning about taking and saving lives. These situations were specifically designed to shed light on tensions in orientations about the value of life (Edmonds, 2013; Foot, 1967; Thomson, 1985). However, past psychological research has not adequately investigated how the trolley car situations bear on people’s moral concerns with the value of life. Moreover, this research has not recognized that the trolley car situations are conflictful because they require a strongly held value – the value of life – to be violated in order to preserve that very value (see Turiel, 2010). Possibly for these reasons, past psychological research has not systematically investigated reasoning about trolley car situations or similar life and death situations among adolescents and adults. Before discussing limitations in past psychological research on responses to trolley car situations, and the claims based on this research, we believe it is essential to consider why philosophers designed the trolley car situations.
Philosophical Origins of the Trolley Car Situations

It was in philosophy, not psychology, that trolley car situations were first formulated and discussed as “trolleyology” (see Edmonds, 2013; Foot, 1967; Sunstein, 2014; Thomson, 1976). The initial variants of the trolley car situations were designed to show that moral considerations include more than counting the sheer number of lives lost and saved. The philosophers who created these hypothetical situations argued that, beyond the number of lives involved, evaluations about sacrificing and saving lives should also be guided by the considerations about whether victims are killed or let die, what rights the persons involved have, and other non-utilitarian considerations (Foot, 1967; Thomson, 1985).

Similarly, the philosopher Bernard Williams argued that utilitarianism is not the only moral principle relevant in life and death situations, for instance in his Critique of Utilitarianism (Williams, 1973). He provided several hypothetical examples of situations that do not have easy solutions because of the conflicts involved. A relevant one pertaining to saving lives involves a man who is confronted with the decision as to whether to shoot one person to his death as a way of saving the lives of 19 people. Part of the situation was described (Williams, 1973, p. 125) as follows: “Jim finds himself in the central square of a small South American town. Tied up against the wall are a row of twenty Indians, most terrified, in front of several armed men in uniform.” As the story goes, Jim learns that the Indigenous people are a random group from a town that experienced acts of protest against the government. The group is about to be killed as a warning to possible protestors. Since Jim is deemed to be an honored guest from another land, he is offered the privilege of killing one of the Indigenous people himself and to mark this as a special occasion, if he does so, the remaining 19 will be set free. What should Jim do? Williams
Moral reasoning 17

asks. The connection to the trolley car situations is obvious. Solely by utilitarian calculations, it
is best to sacrifice one life to save many more.

The main reason Williams constructed these types of situations, as well as the reason
others (e.g., Foot, 1967) had for constructing the trolley car situations, was to illuminate that the
utilitarian calculation is not the only morally relevant consideration. As put by Williams (1973)
regarding the utilitarian answer, “even one who came to think that perhaps that was the answer,
might well wonder whether it was obviously the answer” (p. 125). What other answers could
there be? The other answers or moral considerations to take into account include responsibility
for one’s actions, concerns with one’s integrity, and the psychological effects on the actor: “It is
not that we are utilitarians who are uncertain what utilitarian value to attach to our moral
feelings, but that we are partially at least not utilitarians and cannot regard our moral feelings
merely as objects of utilitarian value” (Williams, 1973, p. 128). These considerations were also
applied in philosophical discourse about different decisions regarding the two versions of trolley
car situations, with some maintaining that considerations other than the number of lives saved, or
the amount of welfare created, are also relevant for moral evaluations (Foot, 1967; Kahane,

Our point here is not that utilitarianism is an unsatisfactory moral philosophy. Rather, we
are noting that the validity of utilitarianism is neither obvious nor widely endorsed by moral
philosophers, and that the trolley car situations were designed to elicit non-utilitarian
considerations (alongside considerations about the number of lives saved). In the switch variant,
many philosophers – like the majority of research participants – have argued that it is permissible
to intervene, sacrificing one person to save five others (Edmonds, 2013; Foot, 1967; Greene et
al., 2001; Thomson, 1985). \textit{Switch} variants of the trolley car situations were originally designed
to show that it is sometimes permissible to sacrifice the life of one person in order to save five others, in line with the utilitarian principle of saving the maximum number of lives (Foot, 1967; Thomson, 1985). However, the structurally different footbridge variant was intended to show that it is not always permissible to sacrifice one person to save five others. Thomson (1976, 1985) proposed that in the footbridge case it was not permissible to intervene. Subsequent empirical work has found that the majority of adults agree with Thomson that it would not be permissible for the bystander to push the large man off the footbridge in order to save the five persons on the track (Edmonds, 2013; Greene et al., 2001).

In the very first discussions about trolley car situations, philosophers proposed several morally relevant features of switch and footbridge type situations as basis for judgments about intervention (Edmonds, 2013; Foot, 1967; Kamm, 2007; Thomson, 1985). Considering some of these proposed features leads to hypotheses about the kinds of considerations adolescents and adults may make in reasoning about trolley car situations. Without aiming to be exhaustive, we briefly describe a few of these proposed considerations below.

**Value of life based on rights.** Several philosophers have argued that human beings are intrinsically valuable and are born with rights to life (Dworkin, 1993; Foot, 1967; Thomson, 1985). From this point of view, all persons involved in the standard trolley car situations have rights to life. However, according to one well-known proposal, it is only in the footbridge situation that the intervention is an infringement on somebody’s rights (Thomson, 1985). That is, Thomson argued that the large man on the bridge had a right not to be pushed, but the single person on the sidetrack did not have a similar right not to have the trolley car redirected.

**Value of life based on numbers.** The right to life is not the only way lives can be considered valuable. Another perspective is to view lives as having additive value, such that the
more lives we can save the better. Most philosophers who have discussed the trolley car situations or related problems share the assumption that the number of lives saved and lost matter for the evaluation of actions. Other things being equal, most argue that it is better to let one person die than to let five persons die (Foot, 1967; Thomson, 1985). This is why most argue that it is permissible to intervene in the switch situation, letting one person die instead of the five workers (though see Thomson, 2008). The controversy surrounding utilitarianism is not about whether consequences, such as the number of lives saved, matter, but whether consequences are the only things that matter for the moral evaluation of actions.

**Involvement of potential victims.** Another aspect discussed as relevant to evaluations about trolley situations is the involvement of the potential victims in the situation. Near the end of her classic article in which she introduced the initial trolley car situation, Foot (1967) noted that “[i]t may also make a difference whether the person about to suffer is one thought of as uninvolved in the threatened disaster” (p. 6). Similarly, Thomson (1985) proposed that it would make a difference to her evaluation of the switch situation if the five persons on the main track were “[m]afia members in workmen's clothing, and they have tied the one workman to the right-hand track in the hope that you would turn the trolley onto him” (p. 1395). According to these views, if it was thought that the persons on the train tracks had accepted that they were in a position of risk (e.g. through a work contract), or even brought about the risk themselves, such involvement may make it more acceptable to sacrifice their lives or let them die.

**Role of intervener: Responsibility for consequences, the natural course of events, and involvement of self.** Another set of related considerations pertain to the involvement of the bystander. In one way or another, many philosophers have argue that there is a morally relevant difference between personally bringing about somebody’s death and somebody or something
else bringing about that person’s death. As expressed by Williams (1973), this is the idea that “each of us is specially responsible for what he does, rather than for what other people do” (p. 99). Foot (1967) brought up this kind of consideration in the context of trolley car situations by differentiating between positive and negative duties. In Foot’s terminology, negative duties referred to the kinds of non-interference we owe to others, whereas positive duties referred to the kinds of help we owe to others. Foot’s argument was that negative duties weigh more heavily than positive duties, such that an obligation not to kill someone (a negative duty) may outweigh the obligation to save someone (a positive duty). That is, other things being equal, it is worse to kill a person than to let a person die, according to this perspective.

A related proposal is that it matters for moral evaluations whether somebody is personally creating a new threat to somebody, for instance by pushing the large person off a footbridge, or merely redirecting a preexisting threat (Edmonds, 2013; Thomson, 1985; Waldmann & Dieterich, 2007). Yet another form of personal involvement considerations is the relation between the potential victims and the intervener. Foot (1967) has hinted at the possibility that familial or other personal relations between the intervener and the potential victims could come with special obligations or rights to help somebody: “[I]t clearly makes a difference whether our positive duty is a strict duty or rather an act of charity: feeding our own children or feeding those in faraway countries” (p. 6). Again, according to this perspective, the intervener’s role in the situation, including his or her relationship to potential victims, is proposed to make a difference for the moral evaluation of his or her action. Indeed, some studies have found that children and adults have obligations toward family members or friends that they do not have toward strangers (e.g. Killen & Turiel, 1998; Miller, Bersoff, & Harwood, 1990).
Thus, from its origins, the philosophical literature on trolley car situations discussed multiple types of considerations as potentially morally relevant: the value of life in terms of rights, the value of life in terms of numbers, the involvement or contractual acceptance of the potential victims, and the responsibility or involvement of the intervener. Philosophers have usually not proposed that any one of these considerations is the sole relevant considerations. Rather, they are so-called *ceteris paribus*, or “other things being equal,” principles, which yield unequivocal answers when two situations differ only in one respect and are equal in all other respects (Foot, 1967; Thomson, 1985). In contrast, such *ceteris paribus* principles do not by themselves indicate how each principle should be weighted relative to other principles. Thus, if a given situation is seen as relating to multiple principles, the individual *ceteris paribus* principles alone do not determine an evaluation. For instance, in the footbridge situation, the two alternatives (push the man vs. do not intervene) differ both in the number of lives lost (one vs. five) and in the involvement of the intervener (causing death vs. letting die). A person who is concerned with both the value of life in terms of numbers and the particular responsibility for one’s own actions will still need to decide which consideration to give priority, which leaves room for conflict and ambiguity.

The philosophical background of the trolley car situations informed our hypotheses about how adolescents and adults would reason about trolley car situations. First, we expected that adults, and possibly adolescents, would raise several different considerations implicated in the situations. Second, we expected that adults, and possibly adolescents, would show indications of conflict as they attempted to coordinate competing considerations about the trolley car situations (e.g. the value of life based on rights and the value of life based on numbers). Third, we expected that evaluations and reasoning would change in systematic ways in response to manipulations of
key features of the trolley car situations. For instance, we expected that if the number of people surviving would be the same whether a person intervened or not, most people would say intervention was wrong and base this evaluation on considerations about the involvement of the intervener. Past research has not investigated the development of these distinct considerations regarding sacrificing and saving lives. Therefore, as noted earlier, we did not have specific expectations about whether adolescents would show sensitivity to all the considerations brought up by adults, and whether there would be developmental differences in which considerations participants emphasized.

The philosophical work on trolley car situations and the moral developmental research discussed above emphasize connections between moral decisions and reasoning. (Whatever philosophers think of what non-philosophers are doing, they presumably view themselves as engaged in reasoning.) In contrast, over the past two decades, psychological research on adults’ responses to trolley car situations have been taken to support a very different set of propositions.

**Dual Process and Intuitionist Approaches to Moral Evaluations**

A substantial body of research with adults has been taken to support the proposition that many, or most, moral decisions are not based on reasoning but on automatic, affective, or unconscious reactions (Hauser et al., 2007; Paxton & Greene, 2010; Wheatley & Haidt, 2005). One of the main lines of research leading to those conclusions involved responses to trolley car situations (Greene et al., 2001).

The findings that have been used to support the propositions regarding the minimal role of reasoning in morality are that (a) in the switch situation, most people say it is permissible to intervene, and (b) in the footbridge situation, most people say it is not permissible to intervene (Greene et al., 2009, 2001; Hauser et al., 2007). Psychologists who have conducted these studies
have interpreted this pattern to show that evaluations about trolley car situations are often “rapid, intuitive, and made with a high degree of certitude” (Mikhail, 2007, p. 144). According to one view, affective and non-reasoned evaluations are most common in response to the footbridge case because people are said to have an affective aversion to pushing someone (Greene, 2014; Greene et al., 2009). In contrast, it is said that people reason by applying the utilitarian principle of maximizing the number of lives saved in the switch variant. Greene (2014, p. 699) summarizes this view as follows: “Characteristically deontological judgments are preferentially supported by automatic emotional responses, while characteristically consequentialist judgments are preferentially supported by conscious reasoning and allied processes of cognitive control.” (As noted by Greene, this use of the terms “deontological” and “consequentialist” are not in accord with most philosophical treatments, hence the qualifier “characteristically.”) According to this account, when people endorse and apply non-utilitarian considerations, for instance about intrinsic rights, they “are not, for the most part, actually engaged in moral reasoning” (Greene, 2014, p. 718). Proposing a similar view, Cushman (2017, pp. 262–263) describes non-utilitarian evaluations as “flops” and “errors” that result from people’s reliance on intuitions.

Several types of evidence have been used to support the view that evaluations about footbridge situations are typically based on automatic and affective reactions, whereas evaluations about switch situations are typically based on utilitarian reasoning (Greene, 2014). We focus on two such bodies of evidence: One regarding the involvement of emotions in evaluations, and the other regarding people’s ability to justify their evaluations about trolley car situations.

**Emotional engagement in evaluations about footbridge situations.** The claim that the footbridge situation, but not the switch situation, elicits emotional reactions is based on several
lines of research. Neuroimaging studies have found greater activation in brain areas typically associated with emotional arousal when people are making evaluations about footbridge-like situations than when they make evaluations about switch-like situations (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene et al., 2001). Moreover, patients with damage to some areas thought to be involved in emotional processing, and persons who score higher on antisocial personality traits or lower on empathic responsiveness, are more likely to say it is permissible to kill one person to save a greater number, for instance in the footbridge situation (Bartels & Pizarro, 2011; Conway & Gawronski, 2013; Gleichgerrcht & Young, 2013; Koenigs et al., 2007). These findings have been interpreted to show that people think it is wrong to intervene in the footbridge situation because of automatic, emotional reactions to the idea of personally causing harm to somebody else. Based on this interpretation, some recent research has taken the presence of non-utilitarian evaluations to indicate that participants were emotionally aroused without separately assessing emotional arousal (Hayakawa, Tannenbaum, Costa, Corey, & Keysar, 2017).

This interpretation assumes a qualitative difference between the emotional processes involved in footbridge evaluations and the (non-emotional) processes involved in switch evaluations. However, the empirical findings do not reveal this qualitative difference. That is, the above-cited studies, and other related studies, do not show that people are emotionally unengaged when making evaluations about switch situations, nor that people do not think carefully about footbridge situations. Indeed, others studies have found participants show heightened emotional arousal even to switch-like situations relative to a baseline condition (Skulmowski, Bunge, Kaspar, & Pipa, 2014). At most, what past studies of emotional involvement in trolley car situations could demonstrate is that people are more emotionally
aroused in response to the footbridge situation than in response to the switch situation. Similarly, relative differences in reaction time does not indicate that the faster response is automatic and emotional, since there is no known temporal cut-off for determining which responses are so fast that they must be automatic. Hence, relative differences in arousal or response times do not reveal that qualitatively different processes (e.g. automatic reactions vs. reasoning) were involved in judging the footbridge and switch situations.

A related problem is that research on emotional responses to trolley car situations often do not assess what people are emotional about. In general, the mere co-occurrence between emotional arousal and evaluations does not show that the evaluations are based on the emotional reaction, since arousal may also follow from conflicts about evaluations. In fact, some research suggests that participants evaluating footbridge-type situations are in part emotionally aroused about the conflictful decision, not just the thought of pushing someone off a bridge (Greene et al., 2004; Kahane, 2015; Moretto, Làdavas, Mattioli, & di Pellegrino, 2009).

Finally, the notion that people become emotionally involved about non-utilitarian moral concerns but not utilitarian ones is inconsistent with standard theories of emotion, which hold that any concern can give rise to emotional reactions (Frijda, 1986; Gross, Sheppes, & Urry, 2011; Lazarus, 1991). Conversely, the mere presence of emotion, or relative differences in emotional arousal, does not indicate an absence of reasoning or other cognitive processes.

In short, a number of empirical and theoretical considerations challenge the notion that non-utilitarian evaluations are based on emotional reactions, while utilitarian evaluations are based on non-emotional considerations. Instead, these considerations indicate that emotion and thought are intertwined, both in the sense that emotions involve cognitive processes and in the sense that people can be emotional about any issue they reason about (provided they care enough
about the issue). We return to these contrasting perspectives on the roles of emotions in morality in the General Discussion.

**Justifications about trolley car situations.** In determining whether evaluations are based on automatic reactions, some of the most crucial evidence concerns people’s ability to justify their evaluations. It has been claimed that people are sometimes unable to justify moral evaluations, a phenomenon sometimes referred to as “moral dumbfounding” (Haidt et al., 2000). In an unpublished study viewed as the main demonstration of moral dumbfounding, the researchers reported that participants would sometimes say that an action was wrong without, upon repeated probing, being able to explain why (Haidt et al., 2000). People’s inability to justify moral evaluations would support the view that moral evaluations are not based on reasoning, but rather on some form of intuition as defined by Haidt et al., the sources of which are consciously inaccessible (Hauser et al., 2007). (See Bruner [1960] and Shweder, Turiel, and Much [1981] for different perspectives on the types of reasoning involved in intuitions, and Jacobson [2012] and Royzman, Kim, and Leeman [2015] for critiques of evidence for moral dumbfounding.)

People’s ability to justify evaluations about trolley car situations have been directly investigated in two studies (Cushman et al., 2006; Hauser et al., 2007). In these studies, participants were presented with several variants of trolley situations and asked to judge whether intervention was permissible. Participants were subsequently asked to explain differences in their evaluations, for instance why they judged intervention as permissible in the switch situation but not the footbridge situation. Researchers then decided whether participants provided a sufficient explanation of difference in evaluation (i.e. identified a “factual” difference), provided an insufficient explanation, or added assumptions not in the initial description of the situation.
In one study (Hauser et al., 2007), 17 percent of participants were deemed to provide “sufficient” justifications for why they evaluated the situations differently. Forty-five percent of participants provided justifications that were discounted because they were deemed to include “added assumptions” not included in the situation descriptions read by participants. Data in the study by Cushman et al. (2006) were reported at the level of individual justifications, not at the level of individuals. They reported that 30 to 80 percent of justifications were deemed “sufficient” for explaining why two situations were evaluated differently. These findings were taken to show that (at least some) evaluations about footbridge situations result from supposed automatic unconscious reactions rather than reasoning.

These two studies of justifications about trolley situations have several limitations. One is that participants were only asked to justify the differences between their evaluations in switch and footbridge situations, and not to justify the initial evaluations. That is, participants were asked why they had judged or rated the footbridge and switch situations differently, but not why they initially made those evaluations. This feature increases the memory demands, since participants are asked to retrieve their justifications long after they made their evaluations. In addition, it may have been more challenging for people to explain why they rated two situations differently than to explain their evaluation of each situation separately. It is possible, for instance, that a person reasons while evaluating two situations separately, yet does not reason about the differences between those two situations.

A second limitation is that the coding categories used in these two studies were normative rather than descriptive (Cushman et al., 2006; Hauser et al., 2007). By classifying justifications as “sufficient” or “insufficient,” the researchers made a determination of which justifications were satisfactory using criteria that may not have been shared by participants themselves (or
other researchers). For instance, justifications considered insufficient included cases in which a participant said that in the footbridge variant, but not the switch variant, the protagonist would be *killing* the one person and that killing is wrong. The justification would have been considered insufficient because it allegedly did not point to a factual difference between the switch and footbridge situations. However, whether the statement about killing points to a factual difference depends on what the participant means by “killing.” If the participant meant to say that in the footbridge case, but not in the switch case, the protagonist would be responsible for purposefully and directly inflicting deadly harm on another person (i.e. “killing”), this *would* count as a factual difference between the footbridge and the switch scenarios.

Despite the above limitations, findings from Cushman et al. (2006), and Hauser et al. (2007) have been used to support the claim that people often cannot provide justifications for their moral evaluations. This claim in turn is taken to support the theory that many or most moral evaluations are based on “intuition,” the basis of which is inaccessible to people. As put by Mikhail (2007, p. 144), “individuals typically have difficulty producing compelling justifications for these judgments: thus, trolley-problem intuitions exhibit a dissociation between judgments and justifications (Cushman et al., 2006; Hauser et al., 2007).” Similarly, Haidt (2007, p. 998) noted that “people can sometimes be ‘morally dumbfounded’—they can know intuitively that something is wrong, even when they cannot explain why (Cushman, Young, & Hauser, 2006).” (see also Cushman, Young, & Greene, 2010, p. 47). Hence, research on people’s ability to justify their evaluations about trolley car situations plays a major role in contemporary debates about contrasting psychological approaches to morality and its development.

Clearly, there is a major discrepancy between the developmental findings and the related interpretations mentioned regarding cohesive and complex configurations of moral evaluations and the interpretations stemming from research on adults’ responses to trolley car situations. (See also Haidt’s [2001] propositions regarding emotionally driven moral “intuitions.”) Considering the discrepancy between developmental research and these studies with adults is useful in part because developmental theories have sought to explain how children develop into adults. We may ask, why the discrepancy? Is it that in adulthood morality shifts to non-reasoned, emotionally-driven decisions? That is unlikely, especially since moral reasoning has been documented among adults too (Turiel, 2015).

One reason for the discrepancy between the developmental and trolley car findings is in the structure of the trolley car situations. We noted how much of the developmental research posed straightforward situations involving moral issues (e.g., someone arbitrarily inflicting harm) in order to investigate children’s and adults’ fundamental moral concerns. In contrast, the trolley car situations, as discussed in the philosophical literature, involve conflicting considerations over the strongly held value of preserving lives. Put in terms used by Foot, “conflicts can occur between negative duties (such as the duty not to kill people) and positive ones (such as the duty to save people)” (Sunstein, 2014, p. 14).

Moreover, as discussed above, the empirical research on responses to trolley situations has not adequately examined reasoning around choices as to which is the best course of action. This research has also not investigated whether people perceive conflicts between competing yet legitimate moral considerations in the trolley car situations. Instead, the differences obtained
between evaluations about variants like the switch and footbridge have been taken to signify that reasoning is only involved when people are counting the number of lives saved, and not when judging on the basis of non-utilitarian considerations such as intrinsic moral rights and responsibilities.

The research presented in this Monograph was based on the proposition that trolley car situations involve reasoning about not only the number of lives saved but also about a number of other morally relevant aspects of the situations. Furthermore, we expected that reasoning would be involved in decisions about trolley car situations, and that evaluations would be associated with different perceived situational features.

**What kind of evidence is needed to establish whether people reason about trolley car situations?** In empirically investigating the contrasting claims about the role of reasoning in moral evaluations, it is necessary to consider what would constitute evidence of moral reasoning.

By saying that moral evaluations are based on reasoning we mean that evaluations are formed on the basis of considerations that people can articulate and endorse. The considerations involve reasoning in the sense that they are applied consistently across situations. For instance, moral concerns with the value of life are seen across a number of situations involving diverse issues such as physical harm, murder, abortion, and euthanasia (Dworkin, 1993; Turiel & Killen, 2010). Moreover, moral reasoning implies that people have some level of awareness of their moral considerations, and can typically articulate and endorse them (except in unusual cases, when stress, memory demands, or other factors could make reasons difficult to retrieve, see e.g. Jacobson, 2012). Thus, in the present research, we expected that participants would provide justifications for why they thought it was wrong to push the person off the footbridge.
Importantly, when we say “moral reasoning” we do not mean “conscious moral reasoning.” That is, we are not assuming that individuals always go through some form of explicit internal dialogue before forming an evaluation based on reasoning. There are two reasons why we avoid viewing conscious, explicit steps of thinking as criterial for reasoning. First, most general views of reasoning imply that reasoning can happen very quickly and effortlessly (Adler & Rips, 2008; Harman, 1986). While reasoning certainly could be conscious and explicit, it does not need to be. Secondly, the definition makes it difficult to determine empirically whether people (consciously) reasoning about moral issues. Perhaps reflecting how difficult it is to operationalize these definitions, researchers who argue that most moral evaluations stem from unconscious, automatic reactions, not from conscious reasoning, do not typically assess whether participants consciously reason before forming evaluations (Haidt & Bjorklund, 2008a; Paxton & Greene, 2010; for related points, see Kihlstrom, 2008).

That individuals apply moral considerations across situations does not imply that they always give priority to these considerations. We discussed earlier how multifaceted situations sometimes require individuals to give priority to some considerations over others. Indeed, the trolley car situations are designed to make people violate their concern for the value of life, since the protagonist must either sacrifice somebody or let somebody die. However, when giving priority to one consideration over another, we hypothesize that people will remain concerned with the non-prioritized concern, and express conflict about their evaluation. For instance, we expected that participants would indicate that they were conflicted about their evaluations about both the standard switch and footbridge situations, given that any action in these situations involves the loss of one or more lives.
A final piece of evidence required to show that people engage in moral reasoning is that their evaluations and stated justifications are sensitive to features of the situations under evaluation. That is, if people tend to say it is wrong to push the person off the footbridge to save the five workers because he was innocent and had nothing to do with the situation, we would expect that participants’ evaluations and justifications would change if the person was no longer innocent. For instance, if the footbridge situation were altered so that the person on the footbridge had set off the runaway trolley car in order to kill the five workers, participants would expectably be more likely to view it as permissible to push the person off the footbridge and likely to reference the person’s lack of innocence in justifying these evaluations (see Study 2).

Thus, there are two main questions for determining whether people are reasoning about trolley car situations. The first question is whether people articulate justifications for their evaluations. The second question is whether the concerns referenced in their justifications are applied across situations, as reflected in expressed conflicts about multifaceted situations and in altered evaluations and justifications when situations change in relevant ways.

The Present Research

The present research was guided by the proposition that moral reasoning about sacrificing and saving lives can involve multiple moral considerations about the value of life, not only utilitarian ones. In response to multifaceted situations, such as the trolley car situations, we proposed that individuals often seek to coordinate conflicting considerations. This research examined how individuals’ explained their evaluations about trolley car situations, including whether they weighed different considerations against each other. Furthermore, we examined
whether people’s evaluations and justifications were sensitive to variations in situational features.

Most of the research with trolley situations has been conducted via electronic surveys with adults (often undergraduate students) during which participants indicate one of several predefined options (such as “okay” or “not okay”). In the few studies that have elicited explanations, participants are asked to explain their reasoning by typing into a text-box in response to a single question (e.g. Cushman et al., 2006). The design of Study 1 differed from past studies in its use of semi-structured interviews, its inclusion of non-undergraduate student samples, and its combination of in-person and online interviews.

Semi-structured interviews are a common method of studying evaluations and reasons or justifications in children and adults, entailing a series of standard questions aimed at uncovering salient features of their thinking (Damon, 1977; Kohlberg, 1963; Piaget, 1929; Turiel, 1983). The evaluations and justifications provided are subsequently analyzed using detailed coding systems. This method has provided data on configurations of thought in children, adolescents, and adults regarding moral and other social issues.

Another central feature of the present research was the investigation of age differences in responses to trolley car situations in adolescents and adults. It is important to study evaluations about life and death situations in individuals prior to college age. Moral issues presented to young children largely involve wellbeing, property rights, and fairness rather than situations about taking lives. However, as noted, it is both feasible and important to study evaluations about life and death among adolescents. On the one hand, it appears that adolescents do make evaluations about an intrinsic right to life that should not be violated (Colby & Kohlberg, 1987a, 1987b). This would suggest some degree of continuity in responses to trolley car situations from
adolescence to adulthood. At the same time, we noted how age differences have sometimes emerged in adolescence in response multifaceted situations that require coordination of competing considerations (Helwig, 1995; Nucci & Turiel, 2009). Thus, we did not know prior to conducting this research whether developmental continuities or developmental differences would be most prominent in responses to life and death situations from adolescence to adulthood.

The two studies reported here were designed to address the following three questions regarding adolescents’ and adults’ responses to trolley car situations. (We provide additional details on how each question was addressed in Chapters 2 and 3.)

(1) How do adolescents and adults justify their evaluations about situations involving sacrificing and saving lives? Based on past developmental research, as well as our review of the philosophical literature, we predicted that adolescents and adults would reference a number of distinct considerations in support of their evaluations about the standard trolley car situations. These considerations included the value of life based on rights, the value of life based on numbers, the involvement of the potential victims, and the involvement of the potential intervener. In contrast, the dual-process view would predict that participants would primarily reference the number of lives saved, based on the proposal that moral reasoning mainly involves utilitarian considerations. This question was primarily addressed by Study 1, although Study 2 also investigated participants’ justifications about variants of trolley car situations.

(2) Are adolescents and adults conflicted about their evaluations about the standard trolley car situation? Since we expected participants to be concerned with multiple, competing considerations in the standard trolley car situations, we expected them to be conflicted about their evaluations. Conflicts were assessed by coding whether people provided justifications for the opposing evaluation, rejected the consequences of the chosen action as unacceptable, and
accepted that others would judge the situations differently. We expected such conflicts to be common in both switch and footbridge situations, since we expected both to involve competing considerations. In contrast, the dual-process view would predict that participants may be conflicted about the footbridge situation, but not about the switch situation (Greene, 2008; Greene et al., 2004, see next chapter). This question was addressed by Study 1.

(3) Are evaluations and justifications sensitive to variations in key features of the situations? By proposing that people’s evaluations are based on moral considerations, we hypothesized that those considerations would be applied across situations. Hence, when key features of the situations changed, we expected participants’ evaluations and justifications to change correspondingly. For instance, if participants’ discussed the value of life in terms of intrinsic rights when explaining why it was wrong to push the person off the footbridge in the standard variant, we expected them to accept interventions if the person pushed off the footbridge would only be scratched and bruised, not killed. In contrast, the dual process view have proposed that people’s justifications are, with the exception of utilitarian justifications, post hoc rationalizations that do not typically play a role in people’s evaluations. From this point of view, there should be little correspondence between the justifications people provide about one situation and the evaluations they provide about a different situation. Study 1 compared patterns of evaluations and justifications across one standard and four non-standard variants of the switch and footbridge situations. Study 2 further explored this question by manipulating the involvement of the potential victims in the trolley car situations, following up on findings from Study 1.
Chapter 2: Study 1: Investigating both Evaluations and Reasoning about Systematically Varied Trolley Car Situations

The goal of Study 1 was to investigate adolescents’ and adults’ evaluations and justifications about several variants of trolley car situations. Using structured interviews with several questions about each situation and systematic coding of justifications, we investigated the kinds of considerations adolescents and adults brought to bear on trolley car situations and their expression of conflict or difficulty in evaluating trolley car situations. We expected that participants at all ages would invoke several distinct types of considerations, not just utilitarian considerations about the number of lives saved, and that they would sometimes be conflicted when seeking to coordinate competing considerations about sacrificing and saving lives.

Study 1 included adolescents recruited from a high school, undergraduate students recruited from a research university, and a sample from the wider adult population recruited via Amazon Mechanical Turk (mTurk). To assess whether responses to online surveys differed from responses to in-person interviews, we included one sample of undergraduate students interviewed in person and one sample interviewed via an online survey. (Because of the requirement of parental consent, the adolescent sample was also interviewed in person.)

Study 1 used semi-structured interviews conducted in-person and online to investigate the following three questions:

(1) **How do adolescents and adults reason about trolley car situations?** The study investigated people’s justifications for their evaluations about standard trolley car situations. Whether people hold such justifications is at the crux of some debates in research on moral psychology. If people are often unable to provide justifications for their evaluations, this would indicate that the evaluations are not based on reasoning. In contrast, if people do justify their
evaluations in coherent ways, as we hypothesized, this would support the proposition that evaluations about trolley situations are based on reasoning.

(2) Are people conflicted about their evaluations of trolley car situations? We propose that responses to trolley car situations involve balancing of conflicting moral concepts. That is, we expected that even when participants viewed intervention as wrong (e.g., letting the five persons die instead of sacrificing the one person), participants would be sensitive to the considerations in favor of the opposing evaluation (Białek & Neys, 2016). Therefore, we hypothesized that people’s responses to the standard switch and footbridge situations would entail conflict.

Our expectations that most people will be conflicted about the switch and footbridge situations contrast with the viewpoint that people typically respond to the footbridge situation with an automatic, emotional reaction (Greene, 2013, 2014). Based on reaction time and neuroscientific data, Greene and colleagues (2004) argued that “[t]he footbridge and infanticide [a mother killing her unwanted infant] dilemmas tended to be easy” (p. 395), meaning that judging these situations tended to involve “no significant cognitive conflict and no need for extended reasoning or cognitive control” (p. 391). Conversely, the dual process model hypothesizes that the switch situation elicits little to no automatic and emotional reactions. Hence, participants are said to evaluate the switch situation by counting the number of lives lost and saved, again without experiencing conflict: “[T]here is, according to this model, no emotional response (or much less of one) to override in such cases” (Greene, 2008, p. 44). We note that the dual process account does expect conflict to occur for the minority of participants who are presumed to override their automatic, affective reactions, for instance by judging that it is permissible to push the person off the footbridge (see Greene et al., 2004, 2001). Accordingly,
people are said to effortlessly judge that it is wrong to push the person off the footbridge (although conflict can occur when people decide to override their aversion to pushing). Evaluations about the switch situation are also typically portrayed as straightforward, although for different reasons. Evaluations are hypothesized to be based merely on counting the number of lives lost and saved. Hence, the dual process explanations of the typical responses to the footbridge and switch situations (i.e. intervention as permissible due to utilitarian calculus in the switch situation, but not in the emotionally laden footbridge situation) imply that these evaluations are made with little or no conflict.

By contrast, if people take into account a number of partly conflicting evaluative considerations as they make evaluations about trolley car situations, we would expect that they will express being conflicted about their evaluations regardless of their decision. In the present study, we investigated such conflict by assessing whether people explicitly considered arguments for the opposite evaluation, negatively evaluated the consequences of their favored option, and accepted alternative evaluations.

(3) Are people’s evaluations and justifications sensitive to morally relevant variations of the situations? We hypothesized that there are different considerations involved in people’s reasoning, which influence their evaluations. Thus, when situations change in ways relevant to those considerations, we expected the evaluations and justifications to change accordingly. These questions were addressed in Study 1 through variations in the trolley car situations presented to participants. This study included the standard switch and footbridge situations, as well as the following four pairs designed to investigate the role of key considerations in people’s evaluations.
Changing characteristics of the consequences of being pushed. It has been maintained that people have an emotional aversion to intentionally pushing the person off the bridge because they have an aversion to the intentional use of “personal force,” or acts in which “the agent directly impacts the victim with the force of his/her muscles” (Greene, 2014, p. 709; Greene et al., 2009). When imagining the act of pushing, people are said to have an aversive reaction that leads them to judge the act as impermissible. We propose that it is not the act of pushing that people view as central, but the act of directly causing someone’s death. Thus, we hypothesized that if the person being pushed off the bridge was only scratched and bruised by stopping the train, most participants would find it permissible to push him, even though the protagonist would still be using intentional personal force toward someone else. For these purposes we used a situation referred to as 1 scratched vs. 5 killed.

Changing characteristics of the people saved. We propose that the negative evaluation of pushing is based on balancing the wrongness of taking a life against one’s responsibility for the lives of the five persons on the track (see Williams, 1973). We therefore hypothesized that people would be more likely to judge that it is permissible to push the a stranger off the bridge if they would be saving their sister and her four children – involving an evaluation that one is responsible for the lives of family members (referred to as 1 killed vs. 5 family killed).

The role of the value of human life. We also expected that the negative evaluation of pushing is based on a respect specifically for human life. Although people would expectably be reluctant to push an animal off a bridge for no reason, we hypothesized that they would be willing to do so in order to save a human life. Hence, we hypothesized that people would be more likely to find it permissible to push a dog off the bridge, or shift the train toward a dog, in order to save the five persons than to sacrifice a human life (1 dog killed vs. 5 people killed).
Distinctions between intervention and non-intervention when consequences are the same. Several accounts of the trolley situations suggest that people treat the switch situation purely in terms of the number of people who die. If so, people should not have a negative evaluation about switching the train toward one person to save one other person on the main track. Only one life is lost either way, making the outcomes equivalent from the point of view of utilitarian reasoning. In contrast, we hypothesized that participants would find intervention more objectionable than non-intervention other things being equal, and hence that they would be less likely to find intervention permissible in situations involving one life for one life (1 killed vs. 1 killed) than in the standard (1 killed vs. 5 killed) situations.

In summary, Study 1 investigated responses to the standard switch and footbridge variants of the trolley car situations as well as four new variants. The study investigated both how (if at all) adolescents and adults reasoned about the different situations involving sacrificing and saving lives, as well as developmental continuities and changes in responses to such situations. In doing so, the study addressed both fundamental questions about the development of moral concepts regarding the value of life, and current debates between reasoning-based and dual-process approaches to moral psychology.

Method

Participants. Study 1 had a total of 288 participants (59% female): 48 adolescents participating in an in-person interview (\(M_{age} = 16.0\) years, \(SD_{age} = 0.6\) years, 50% female, 50% male), 48 undergraduate students participating in an in-person interview (\(M_{age} = 22.0\) years, \(SD_{age} = 3.9\) years, 69% female, 31% male), 96 undergraduate students participating in an online interview (\(M_{age} = 20.9\) years, \(SD_{age} = 2.6\) years, 76% female, 24% male), and 96 adults recruited via mTurk for an online interview (\(M_{age} = 35.6\) years, \(SD_{age} = 11.4\) years, 43% female, 57%
male). There was slight overlap in the ages between the mTurk and college student samples, such that the lower 10 percent of the ages in the mTurk sample overlapped with the upper 10 percent of the ages in the college student samples.

Adolescents were recruited from a high school in suburban New England. The adolescent sample was 73 percent White Non-Hispanic, 13 percent Asian-American, and 14 percent other ethnicities. The children at the school are predominantly from primarily middle and upper-middle class backgrounds and most children have at least one parent with a college degree.

Undergraduate students were recruited from a research participation pool at a large public research university in the Western United States. Due to an administrative error, ethnicity information was not collected for college students interviewed in person. The ethnicity distribution of the undergraduate population of the university as a whole is: 44 percent Asian/Pacific, 32 percent White Non-Hispanic, 14 percent Hispanic, and 10 percent other. The undergraduate sample participating in online interviews was 56 percent Asian American, 17 percent White non-Hispanic, 15 percent Hispanic, and 12 percent other. Most undergraduate students at the university are from middle or upper-middle class families and have at least one parent with a college degree. Undergraduate students received course credit for participation.

Participants recruited via mTurk were 80 percent White Non-Hispanic, 9 percent Asian American, and 11 percent other ethnicities. Participants had IP addresses based in the United States. In a recent study of demographics of mTurk participants, 45 percent had college degrees and the average individual income was $54,000 (Levay, Freese, & Druckman, 2016), which is within the range of what is typically considered middle-class income in the United States (Thompson, Hickey, & Thompson, 2016). mTurk participants received $2.25 for their participation.
Prior to participation, all participants signed written assent forms (adolescent participants), written consent forms (undergraduate in-person participants), or electronic signature indicating consent (online participants).

**Materials.** The wording of descriptions of the situation was based on past psychological research on trolley car situations (Cushman et al., 2006; Greene et al., 2009, 2001; Hauser et al., 2007). One footbridge variant and one switch variant was created for each of the five situation types (standard [1 killed vs. 5 killed], 1 scratched vs. 5 killed, 1 killed vs. 5 family killed, 1 dog killed vs. 5 people killed, and 1 killed vs. 1 killed). The standard (1 killed vs. 5 killed) switch situation was described as follows:

You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a switch. The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate track where there is a single railway workman. The single workman will be killed if you flip the switch.

The standard (1 killed vs. 5 killed) footbridge situation was described as follows:

You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a large stranger. The only way to save the five workmen is to push the stranger, who would fall onto tracks and stop the train. The stranger will be killed if you push him.

The descriptions of non-standard variants of switch and footbridge were identical to the standard variants except for the following: In the 1 scratched vs. 5 killed situations, the one person would be “scratched and bruised” instead of “killed” if the protagonist intervened. In the
1 killed vs. 5 family killed situations, the five persons on the track were described as “your sister and her four children” instead of “five workmen.” In the 1 dog killed vs. 5 people killed situations, the single individual potentially sacrificed for the five workers was a dog instead of a human being. Finally, in the 1 killed vs. 1 killed situations, only one worker was on the main track instead of five. The complete texts of the non-standard situation descriptions were as follows:

1 scratched vs. five killed situations

Switch variant. You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a switch. The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate track where there is a single railway workman. The single workman will be scratched and bruised if you flip the switch.

Footbridge variant. You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a large stranger. The only way to save the five workmen is to push the stranger, who would fall onto tracks and stop the train. The stranger will be scratched and bruised if you push him.

1 killed vs. 5 family killed

Switch variant. You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are your sister and her four children, who will surely be killed unless you do something. Next to you there is a switch. The only
way to save your sister and her four children is to flip the switch, which would divert the train onto an alternate track where there is a single railway workman. The single workman will be killed if you flip the switch.

*Footbridge variant.* You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving down the track. In the path of the train are your sister and her four children, who will surely be killed unless you do something. Next to you there is a large stranger. The only way to save your sister and her four children is to push the stranger, who would fall onto tracks and stop the train. The stranger will be killed if you push him.

*1 dog killed vs. 5 people killed*

*Switch variant.* You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a switch. The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate track where there is a large dog. The dog will be killed if you flip the switch.

*Footbridge variant.* You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a large dog. The only way to save the five workmen is to push the dog, who would fall onto tracks and stop the train. The dog will be killed if you push him.

*1 killed vs. 1 killed*

*Switch variant.* You are standing near a train track as you see a runaway train car moving down the track. In the path of the train is a railway workman, who will surely be
killed unless you do something. Next to you there is a switch. The only way to save the workman is to flip the switch, which would divert the train onto an alternate track where there is another railway workman. The latter workman will be killed if you flip the switch.

*Footbridge variant.* You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving down the track. In the path of the train is a railway workman, who will surely be killed unless you do something. Next to you there is a large stranger. The only way to save the workman is to push the stranger, who would fall onto tracks and stop the train. The stranger will be killed if you push him.

**Procedures.** Procedures for in-person and online interviews were similar, except that in-person interviews were conducted by a researcher, who read the situation descriptions and asked the interview questions, whereas the online interviews were administered via the survey platform Qualtrics. In-person interviews with adolescents and college students were conducted in a separate room at the university or school of the participant. In-person interviews typically lasted between 20 and 40 minutes. The online participants read the situations and responded to the same interview questions as the in-person participants, except that the interviewer in the in-person interviews was able to ask questions of clarification.

Each participant was first interviewed about the standard pair of situations (switch and footbridge). Then, participants were randomly assigned to receive two non-standard situation pairs. The order of switch and footbridge was counterbalanced across participants, but constant across the situation pairs for a given participant. The order of the two non-standard situation pairs was also counterbalanced across participants. For example, a given participant might receive the following sequence of situations: standard (1 killed vs. 5 killed) switch, standard (1
Moral reasoning

killed vs. 5 killed) footbridge, 1 killed vs. 5 family killed switch, 1 killed vs. 5 family killed footbridge, 1 killed vs. 1 killed switch, and 1 killed vs. 1 killed footbridge. This design yielded 24 unique sequences of situations.

For each situation, participants were asked the following questions (wordings were modified as appropriate for non-standard situations): (a) Is it alright or not alright to [push the stranger/flip the switch]? (b) Why/Why not? (c) Why is it better to [intervene/not intervene]? (d) Why is it alright to let the [1 person/5 persons] die? (e) Suppose someone else decides to [opposite decision of participant]. Would it be alright or not alright to do that?

After responding to the three situation pairs, participants were reminded of their overall evaluations to standard (1 vs. 5) switch and footbridge situations and asked to compare their perceptions of the two situations. If participants had provided a different evaluation about the two situations (e.g. said intervention was permissible in the switch situation but not in the footbridge situation), they were reminded of it and asked: “Why is there a difference between the two situations?” If participants had provided the same evaluation about the two, they were asked: “Do you think there is a difference between the situations? Why/Why not?” We only report analyses of these situation comparisons for participants who showed the typical pattern of evaluations, deeming intervention to be permissible in the switch situation but not in the footbridge situation (see Results).

**Coding and data analysis.** The answer to question (a) provided the participants’ overall evaluation of the situation. Justifications for and against the overall evaluation were coded for question (b), (c), and (d). Participants’ conflicts about their evaluations were assessed by analyzing whether they provided justifications favoring the opposite evaluation (b, c, d), refused to accept the consequences of their favored course of action (d) (e.g. “But it’s not alright for the
one person to die!”) and acceptance of another person judging the situation differently (e) (e.g. a person saying it was alright to intervene when the participant said it was not alright to intervene).

Evaluations were coded as either “alright” or “not alright.” Coding categories for justifications, which are listed in Table 1, were developed from a review of a subset of interviews. An additional coding category for “don’t know” responses was included in the coding system, but was used so infrequently (2% of justifications) that the data were not analyzed. (All participants coded as saying “don’t know” still provided at least one other justification, suggesting that the “I don’t know” may have referred to the difficulty of making an evaluation rather than the difficulty of justifying that evaluation.) For each question in each interview, coders indicated whether a given justification type was used and, if used, whether the justification was used in favor of or in opposition to the person’s evaluation. Twenty percent of interviews were randomly selected for double-coding in order to assess inter-coder agreement. Agreement for evaluations, as measured by Cohen’s κ, was κ = .97. Average agreement for justifications was κ = .83 (range: .75 to .88).

Evaluation and justification data were analyzed using Generalized Linear Mixed Models (GLMMs) with a logistic link function and binomial error distribution (Fox, 2008; Hox, 2010). Data were analyzed in R (version 3.4.0) and GLMMs were fitted using the lme4 package. For evaluations, models were fitted to the dichotomous variable indicating whether the participant responded “not alright” to a given question. Data for each justification were analyzed separately. The dependent variable was a dichotomous variable indicating whether a given justification was
used. Models included random intercepts for participants and, unless otherwise noted, fixed effects of switch vs. footbridge, situation pair (standard, 1 scratched vs. 5 killed, 1 killed vs. 5 family killed, 1 dog killed vs. 5 people killed, 1 killed vs. 1 killed), and group (adolescent, undergraduate in-person, undergraduate online, mTurk). As preliminary analyses revealed no significant effects of gender or order, these variables were not included in the final models. In the analyses of justifications, models also included a fixed effect of evaluation (alright, not alright). Hypotheses were tested using likelihood ratio tests ($D$, for tests of model fit) and Wald tests ($W$, for tests of individual regression coefficients).

**Results**

**Evaluations. Standard Situations (1 killed vs. 5 killed).** Using the standard switch and footbridge situations, past studies have found that most people say it is permissible to flip the switch to save the five lives but not permissible to push the stranger onto the tracks to save the five. Our findings with the standard situations were consistent with past findings, although our participants were somewhat less likely to say that intervention in the switch scenario was “alright” than participants had been to say that such intervention was “permissible” in some past studies. In the present study, 57 percent of participants said intervention was alright in switch situation, and 16% said intervention was alright in the footbridge situation, binomial GLMM: $D[1] = 146.70, p < .001$.

There was also a significant effect of group on evaluations of alright, $D(3) = 15.73, p = .001$. The undergraduate students in the in-person interviews were significantly more likely to say that the intervention was permissible (50%) than undergraduate students in online interviews (34%), Wald test: $W(1) = 5.43, p = .020$, and adult participants (28%), $W(1) = 9.22, p = .002$, but did not differ from adolescent participants (45%), $W(1) = 0.47, p = .49$. Adolescent participants
in turn differed significantly from adult participants, \( W(1) = 6.23, p = .013 \), but not from undergraduate online participants, \( W(1) = 2.78, p = .095 \).

A common explanation for why people are less accepting of intervention in the footbridge situation than in the switch situation is that people have emotionally-based aversions to pushing others and to using others as means to an end (Cushman et al., 2006; Greene et al., 2008). These aversions, it is argued, prevent people from applying utilitarian principles to the footbridge scenario, which would otherwise favor sacrificing one person to save five.

Non-standard situations in the present study were included to test whether people are generally opposed to pushing others and to using others’ as means (1 scratched vs. 5 killed, 1 killed vs. 5 family killed, 1 killed vs. 1 killed), as well as to relate to the value of human life (1 dog killed vs. 5 people killed). Table 2 lists the percentages of participants saying intervention was permissible as a function of situation and group. The difference between the switch and footbridge situations depended significantly on situation variant (standard, 1 scratched vs. 5 killed, 1 killed vs. 5 family killed, 1 dog killed vs. 5 people killed, 1 killed vs. 1 killed), \( D(4) = 26.95, p < .001 \). We therefore proceeded to analyze the data for each non-standard situation variant separately.

\[ \text{---------------------------} \]

\text{INSERT TABLE 2 HERE} \[ \text{---------------------------} \]

\textit{1 scratched vs. 5 killed: Changing characteristics of the consequences of being pushed.} Our expectation was that people evaluate negatively interventions in the footbridge situation based a concern with not physically causing the death of another. We therefore predicted that participants would be far less likely to judge the intervention as wrong when the person on the
footbridge would only be scratched and bruised (1 scratched vs. 5 killed) than when the person on the footbridge would be killed (standard). Our account contrast with dual-process accounts, which have argued that people negatively evaluate interventions in the footbridge situation based on aversions to physically pushing somebody and to a person as a means to save five others (Greene, 2014; Greene et al., 2009). This dual-process account would not predict that it would make a big difference whether the person on the footbridge would be scratched or killed, since intervention in both situations would include pushing and using somebody as a means to save others. Additionally, we also expected that participants would be more likely to judge intervention as permissible in the 1 scratched vs. 5 killed switch situation than in the standard switch situation.

The large majority of participants judged the interventions as permissible in both 1 scratched vs. 5 killed variants (80% for the footbridge situation and 92% for the switch situation). The differences from the standard variants were significant for both the footbridge (80% vs. 16%), $D(1) = 171.86, p < .001$, and the switch situations (92% vs. 57%), $D(1) = 73.56, p < .001$. Although smaller, the difference between the switch and footbridge situations remained significant for the 1 scratched vs. 5 killed situation, $D(1) = 44.21, p < .001$. There was no significant effect of group on evaluations about 1 scratched vs. 5 killed situations, $D(3) = 0.00, p = 1.00$.

1 killed vs. 5 family killed: Changing characteristics of the people saved. Another way we assessed if people have an aversion to pushing another that is closer to the standard situations was to change the characteristics of the people saved by putting it into the context of family members. In these situations the individual had to make a decision to push another or throw a switch (leading to one death) in order to save five family members. With regard to the footbridge
situation, if people always view it as wrong to physically push one person to his death to save five others, the identity of the people being saved should not make a major difference for how they judge an act of pushing. Conversely, if people balance evaluations of the act of pushing against their obligations toward family members, they would expectably be more likely to accept pushing and switching in the 1 killed vs. 5 family killed situations than in the standard trolley car situations.

Consistent with our predictions, participants were significantly more likely to judge it as permissible to intervene in the 1 killed vs. 5 family killed situations than the standard situations both for the footbridge situations (55% vs 16%), \(D(1) = 77.73, p < .001\), and the switch situations (76% vs 57%), \(D(1) = 24.18, p < .001\). The difference between switch and footbridge situations, although smaller, remained significant for the 1 killed vs. 5 family killed situation, \(D(1) = 50.52, p < .001\).

For 1 killed vs. 5 family killed situations, there was also a significant effect of group, \(D(3) = 9.10, p = .028\). Undergraduate students who participated online were significantly more likely to say intervention was alright (73%) than adult participants (58%), \(W(1) = 20.48, p < .001\), and in-person student participants (60%), \(W(1) = 17.05, p < .001\). There were no other significant group differences, \(W_s(1) < 0.47, ps > .49\).

**1 dog killed vs. 5 people killed: The role of the value of human life.** This situation was designed to ascertain if evaluations about human lives are implicated in how the standard and other situations are construed. The large majority judged on the side of human lives over the life of a dog in both situations in that they found it permissible to push the dog (74%) and throw the switch (85%). These evaluations were significantly different from the standard situations (footbridge: 74% versus 16%, \(D(1) = 152.90, p < .001\); switch: 85% versus 57%, \(D(1) = 49.05, p\).
<.001). Still, the difference between switch and footbridge remained significant for the 1 dog killed vs. 5 people killed situations, $D(1) = 23.53, p < .001$. For 1 dog killed vs. 5 people killed situations, there was no significant effect of group, $D(1) = 1.35, p = .72$.

1 killed vs. 1 killed: Distinctions between intervention and non-intervention when consequences are the same. We designed the alternative (non-standard) situations based on the premise that individuals coordinate multiple considerations when they evaluate trolley car situations. These considerations include the value of life and the nature of the interventions involved. If the number of lives lost and saved is the same for intervention and non-intervention (e.g. one killed versus one killed), we would expect that most people would deem it impermissible to intervene, regardless of type of intervention.

It was found that the large majority of participants judged intervention as wrong in both 1 killed vs. 1 killed situations (97% in footbridge and 92% in switch). Participants were significantly less likely to say that intervention was permissible in the 1 killed vs. 1 killed situations than in the standard situations, for both footbridge (3% vs. 16%), $D(1) = 21.68, p < .001$, and switch (8% vs. 57%), $D(1) = 139.72, p < .001$, situations. Although the large majority of participants judged intervention to be wrong, the difference between the switch and footbridge situations remained significant for the 1 killed vs. 1 killed situation, $D(1) = 4.60, p = .03$. Finally, there was no significant effect of group on evaluations about 1 killed vs. 1 killed situations, $D(3) = 3.00, p = .39$.

Justifications of evaluations for standard situations. In this study and unlike most previous studies with trolley car situations, participants were also posed with questions designed to assess their reasons (referred to as justifications) for their evaluations about all of the situations. First, we consider the findings on justifications provided for evaluations in the two
standard situations. Participants were able to provide at least one justification to explain their evaluation \((M = 2.1\) justifications). In addition, participants provided more justifications for evaluations that intervention was *not* alright \((M = 2.20)\) than for evaluations that intervention *was* alright \((M = 1.40)\), Linear Mixed Models: \(D(1) = 99.18\ p < .001\).

There was also a significant effect of group on the number of justifications provided, \(D(1) = 50.05, p < .001\). Adolescents provided more justifications \((M = 2.35)\) for their evaluations than students interviewed in person \((M = 1.95)\), \(W(1) = 8.91, p < .001\), students interviewed online \((M = 1.82)\), \(W(1) = 34.51, p < .001\), and adults \((M = 1.75)\), \(W(1) = 48.70, p < .001\). In turn, students interviewed in person provided more justifications than students interviewed online, \(W(1) = 5.90, p = .015\), and adults, \(W(1) = 12.51, p < .001\). Students interviewed online and adults did not differ significantly, \(W(1) = 1.94, p = .16\). There was no significant effect of situation type, \(D(1) = 2.37, p = .12\).

Participants’ justifications indicate that they were not simply reacting without an ability to explain their reasons for their reactions (as might be proposed in certain theoretical perspectives; see Hauser et al. [2007] for such a perspective). That participants have reasons associated with their evaluations is consistent with our past findings and theoretical propositions (Turiel, 1983, 2008b, 2010). However, of particular importance for current theoretical debates is the type of reasons provided, and what these reasons show as to how participants understood the situations, and whether they understood the situations as different from each other. Table 3 presents the proportions of justifications falling into each category.

================================

**INSERT TABLE 3 HERE**

================================
It was found that the patterns of justifications were associated with the evaluations and, in many cases, features of the situation. We present analyses of the use of justification categories at the level of individuals. That is, we model whether a participant used a given justification at least once as a function of situation (footbridge vs. switch), evaluation (alright vs. not alright), and group. Table 4 shows the justifications provided in standard situations as a function of situation (footbridge vs. switch), evaluation, and group.

Value of life based on numbers. The value of life based on numbers justification referenced the number of lives or amount of welfare involved. For both switch and footbridge situations, value of life based on numbers justifications were far more common when participants said intervention was permissible, $D(1) = 426.05, p < .001$ (Table 4). On average, 98 percent used this category of justifications for “alright” evaluations. In explaining why it was permissible to flip the switch, letting one person die to save five, one participant said: “if you see a situation where five people could be killed compared to one person being killed… It would be a horrible outcome obviously if anyone was killed but it would be ultimately better if only one person had to die than five people” (female, 16 years).

In contrast, only 7 percent provided value of life based on numbers justifications when they gave “not alright” evaluations. When participants used value of life based on numbers justifications to explain why intervention was not permissible, it was typically because they said that they were not willing to compare the worth of human lives (i.e. a refusal to base their evaluation on a comparison of the number of lives saved). For instance, one participant said “I
mean, you would save five people, but lives can’t be measured in numbers, so I don’t think you should do anything” (female, 20 years). Controlling for evaluation, there was no significant effect of situation, $D(1) = 0.29, p = .59$, or group, $D(3) = 7.71, p = .05$.

It appears, then, that consideration of the value of life based on calculations of numbers involved was the main reason for participants’ decisions that it is permissible to throw the switch or push the person. For the most part, decisions that it was not alright to engage in the acts were based on the other types of justifications.

**Value of life based on rights.** Value of life based on rights justifications involved statements that people had rights or intrinsic value, or that the participant did not have the right to decide over the situation. References to value of life based on rights were the most common when participants said intervention was not permissible (71%, vs. 16% for “alright” evaluations), $D(1) = 134.77, p < .001$ (Table 4). For instance, one participant explained that it was not alright to push the stranger onto the train track “because it’s violating someone’s rights to kill them. It doesn’t really matter who you’re saving, it’s just not really okay to kill someone instead of someone else” (male, 16 years). There were no significant effects of situation, $D(1) = 1.74, p = .19$, or group, $D(3) = 3.03, p = .39$, on the use of value of life based on rights justifications.

**Victim involvement.** Victim involvement justifications were references to how some persons had, or had not, placed themselves in a context of risk (e.g. being on the tracks constituted a contractual acceptance of the risks associated with railroads). Victim involvement justifications were significantly more common in support of evaluations that intervention was not permissible (30% vs. 7%), $D(1) = 33.98, p < .001$ (Table 4). Victim involvement justifications were also more common in response to the footbridge situation (32%, vs. 11% in switch), $D(1) = 12.11, p < .001$. This justification type was also more common among adolescents (34%) than
among undergraduates interviewed online (17%), \( W(1) = 9.93, p = .002 \) and adults (20%), \( W(1) = 7.98, p = .005 \), but did not differ significantly from undergraduates interviewed in person, \( W(1) = 2.60, p = .11 \). There were no other significant group differences, \( W_s(1) < 2.43, ps > .11 \).

To characterize the considerations categorized as *victim involvement* in more detail, we looked at two sub-categories of these justifications: *innocence* justifications and *accepted risk* justifications. *Innocence* justifications referred to how a potential victim was innocent or outside the situation of danger, for instance: “it’s not fair to sacrifice somebody else’s life who’s just an innocent bystander” (female, 24 years). *Accepted risk* justifications explicitly referred to how potential victims had accepted the risk of being in the situation, typically the risk of being on the train tracks, for instance: “The people on the railroad track […] are inherently taking a risk by standing on this railroad track” (male, 15 years). *Innocence* justifications were the most common type of *victim involvement* justifications in the footbridge situation (63% of *victim involvement* justifications). In the switch situation, *accepted risk* justifications were the most common (57% of *victim involvement* justifications).

**Natural course of events.** *Natural course of events* justifications were references to how certain courses of events were natural. Explaining why she thought it was permissible not to intervene in the switch scenario, one participant put it as follows: “If you don’t do anything, that’s what was going to happen anyway” (female, 23 years). Such *natural course* justifications were more common for “not alright” evaluations (39%) than for “alright” evaluations (7%), \( D(1) = 84.27, p < .001 \) (Table 4). When controlling for evaluations, such justifications were significantly more common in the switch situation than in the footbridge situation, \( D(1) = 8.65, p = .003 \). The effect of group was not significant, \( D(3) = 7.59, p = .06 \).
Responsibility for consequences of actions. Responsibility for consequences of actions justifications stated that the actors would be responsible, or not responsible for certain outcomes. This justification captures one of the notions proposed by Williams (1973) regarding non-utilitarian considerations. Participants were more likely to refer to responsibility for specific outcomes when they said intervention was not alright (52%, vs. 5% for “alright” evaluations), $D(1) = 120.65, p < .001$ (Table 4). These justifications were also more common in the footbridge (49%) than switch (21%) situation, $D(1) = 7.14, p = .008$. There was no significant effect of group, $D(3) = 7.38, p = .061$.

Due to past debates about the role of personal force in the footbridge situations (e.g. Greene, 2009; McGuire, Langdon, Coltheart, & Mackenzie, 2009), we also report summary statistics on three subtypes of the responsibility for consequences of actions justifications. Most such justifications (63%) referred to how the intervention would directly cause the death of the one person (worker or person on bridge), for instance: “I don’t think it will be right to push the stranger onto the tracks because I feel like then I would be playing an active role in killing him even though I would save 5 other people” (female, 19 years). Others invoked the use of “personal force,” (16% of responsibility justifications) although such justifications were often entangled with broader notions of responsibility:

It’s worse to force the other person to die than to let those 5 people die because […] [f]or the one guy I would be actively pushing him in front of the train and murdering him. […] [I]t would be more of an active decision for me to murder someone than just …letting the five people die from the train (female, 19 years).
The remainder of justifications in the “responsibility for outcomes” category (21% of justifications) also explicitly referred to responsibility, or lack of responsibility, for outcomes. One person put it this way (Interviewer: I, Participant: P):

I: Why is it alright to let those five people die?
P: Um…it’s not, but […] it’s not really my problem in a way. I want to help, but […] it would be totally unfair of someone to come up to you after and say [that]. […] I am] not trying to be disrespectful, but […] if this happened to me and someone said “Why’d you let those five people die?” I’d feel like that was kind of an inappropriate question to even ask in the first place, I mean it’s not… um… “Why is it alright for you to ask me that question?” I guess is what I would say to them (male, 19 years).

**Consequences to self.** Consequences to self justifications were references to consequences to the protagonist, for instance statements that one could get in trouble or would find it difficult to carry out the action (e.g., pushing). Participants were more likely to refer to consequences to self when reasoning that intervention was not permissible (21%) than when justifying why intervention was alright (8%), $D(1) = 17.37, p < .001$ (Table 4). One adolescent explained why it was not all right to intervene in the footbridge situation as follows:

I don’t think I would be able to know or live with the fact that I pushed one random person onto the train tracks and killed one person, even though it would save five. I would feel terrible about seeing the five people get killed and knowing that I could have done it, but I think I would feel worse for myself causing someone’s death (male, 15 years).
There was also a significant effect of group, $D(1) = 18.74, p < .001$. References to consequences to self were more common among adolescents (28%) than among college students interviewed online (14%), $W(1) = 9.72, p = .002$, or adults (11%), $W(1) = 14.10, p < .001$.

Students interviewed in person also used consequences to self justifications more often (20%) than adults, $W(1) = 6.40, p = .011$, and near-significantly more than students interviewed online, $W(1) = 3.37, p = .066$. There were no other significant differences between groups, $Ws(1) < 1.30, ps > .25$. There was no significant effect of situation, $D(1) = 0.66, p = .42$.

**Justifications of difference between footbridge and switch situations.** In previous studies, it was found that participants think it is permissible to intervene in the switch situation, but do not in the footbridge situation. We refer to this as the typical pattern of evaluation. In the present study, 41% of participants fit this typical pattern (43% said intervention was wrong in both situations, 16% said intervention was alright in both situations, and 1% said intervention was alright in the footbridge but not switch).

When asked to explain why they judged the footbridge and switch situations differently, 8% said that they no longer judged the situations differently. These participants were excluded from the analyses of such explanations.

All participants who maintained the typical pattern of evaluation provided at least one justification for why they evaluated the standard footbridge and switch situations differently. The majority referred to features of both footbridge (85% of participants) and switch (56%) situations. Table 5 shows the justifications participants provided for a typical evaluation pattern. The upper four rows show the proportion of participants referencing features of the footbridge situation, while the lower four rows show the proportion of participants referencing features of the switch situation.
Across the four groups, the most common justifications for the typical evaluation pattern was to reference responsibility for consequences of actions with respect to the footbridge situation. That is, 30% - 74% of participants said that the actors would be responsible for the death of the man if they pushed him off the bridge, but not the death for any persons killed by the train. The following excerpt illustrates this reasoning:

I: Why is it killing him when you’re pushing him, but not when you’re flipping the switch?

P: Flipping the switch is more abstract, not like pushing, so concrete. It’s kind of a sense of how do you think the definition of killing, it’s like killing by your hand or that the train killed him (female, 21 years).

Relatedly, many (7% - 48%) also used victim involvement justifications for the footbridge situation, stating that the person on the bridge was innocent whereas the persons on the track were already in the situation. One participant, exemplifying the latter type of justification said: “To kill that person who isn’t a railroad worker, who has nothing to do with railroads or trains, and this is me to make the decision that his life has to die to save these other strangers? It just didn’t seem right” (male, 25 years).

**Summary: Justifications of evaluations about standard situations.** Participants’ reasons varied with their evaluations and, in many cases, the type of situation they were evaluating. When judging that intervention was permissible, participants nearly always referenced the number of lives saved (value of life based on numbers). In contrast, when they
judged that it was wrong to intervene, participants tended to reference value of life based on rights, victim involvement, natural course of events, responsibility for consequences, or consequences to self. However, victim involvement and responsibility for consequences of actions justifications were especially common for the footbridge situation, whereas natural course of events justifications were more common for the switch situation. Whereas patterns of justifications were generally similar across groups, victim involvement and consequences to self justifications were more common in in-person interviews (adolescents and students) than in online interviews (students and adults). About 40% of participants judged that intervention was permissible in the switch situation but not the footbridge situation. When asked for reasons for this “typical” pattern of evaluations, the vast majority maintained and justified their pattern of evaluation by referencing distinct features of the footbridge and switch situations.

The conflicts in judging the standard situations. Another objective of the study was to assess whether people viewed the standard switch and footbridge situations as straightforward or if they were conflicted about their evaluations. We focus on three possible expressions of such conflict: (1) Do people view some consequences of their favored course of action as unacceptable? This would indicate conflict because participants would say that even though a particular action was alright, the consequences of that action were at the same time not alright. (2) Do people consider reasons for the opposite evaluation? If so, this would indicate conflict as it would show people considering the reasons for the alternative evaluation, suggesting that they saw compelling reasons to act in more than one way (e.g. both for intervening and not intervening). (3) Would people accept that another person could legitimately judge the situation differently? If so, this would indicate conflict because people could imagine valid reasons for judging the situation differently. In sum, if people view the situations as straightforward, the
answer to each question usually would be negative. By contrast, if people view these situations as complex and struggle to judge them, the answer to each question would be positive. Table 6 summarizes these responses.

Unacceptability of consequences. After participants had indicated whether it was alright or not alright to intervene, they were asked why it was alright for one or five people to die. A substantial proportion of participants (20% - 42%) responded that it was not alright to let anyone die, even if that was the better option (Table 6). Participants were about equally likely to make that evaluation in the footbridge situation (26%) and the switch situation (29%), \( D(1) = 2.96, p = .09 \). There were also no significant effects of group, \( D(3) = 6.60, p = .09 \).

The following excerpt exemplifies a conflict about the consequences of one’s own evaluation:

I: Is it alright or not alright to flip the switch and redirect the trolley to the side track?

P: Alright.

I: Why is it alright?

P: Because sounds like it is better to kill only one person instead of five. But it is still not a good thing. […]

I: Why is it alright to let the one person die?

P: It’s not alright, it just… it just in this situation, like, seems like someone either this one person or those five people have to be killed. If that’s that case, seems like only one person could be killed is better than all five of them (female, 25 years).
Consideration of arguments for contrary evaluations. On average, 12% of participants spontaneously considered reasons for a contrary evaluation. However, there was significant variability between groups in their propensity to consider arguments for opposing evaluations, $D(3) = 24.34, p < .001$. As shown in Table 6, participants in the in-person interviews were more likely to consider opposing arguments (adolescents: 19%, students: 25%) than were online participants (students: 9%, adults: 5%). Adolescents differed significantly from students interviewed online, $W(1) = 4.23, p = .040$, and from adults, $W(1) = 10.31, p = .001$. Similarly, students interviewed in person differed significantly from students interviewed online, $W(1) = 9.75, p = .002$, and from adults, $W(1) = 16.80, p = .001$. There were no other significant group differences, $Ws(1) < 2.30, ps > .12$. There was also no significant effect of situation, $D(1) = 0.69, p = .41$.

In the following excerpt, a participant who judged that it was not permissible to intervene in the switch situation gives clear consideration to the opposing viewpoint:

I: Is it alright or not alright to pull the switch and redirect the trolley onto the side track?

P: I guess it depends on my personal opinion. I would say it doesn’t make it any better. That person still has intrinsic worth but coming from a utilitarian standpoint, it might make more sense to kill one person than five persons but I believe that each individual has their own intrinsic worth and so it’s not morally right to kill them, any of them (male, 20 years).

Table 7 presents the proportion of participants who provided opposing justifications (among participants who provided at least one opposing justification). These justifications are consistent with the data presented above on the justifications provided in favor of each evaluation. For instance, among participants who said it was not alright to intervene in the footbridge and switch situations, the most common opposing consideration was value of life.
Based on numbers. Indeed, the excerpt provided above shows a participant who primarily reasoned in terms of the intrinsic value of human life (value of life based on rights), yet acknowledged the utilitarian (value of life based on numbers) perspective.

Evaluations about opposite decision. When participants were asked whether it would be alright if someone else judged the situation differently, 57% percent of participants said “yes.” There was no difference between the footbridge and switch situations (both 57%), $D(1) = 0.01, p = .91$, nor was there a significant effect of group, $D(3) = 4.76, p = .19$ (Table 6).

Overall expression of conflict. Seventy-one percent of participants expressed at least one of the above indications of conflict (not accepting consequences, considering alternative viewpoint, accepting opposite decisions). The tendency to express difficulty in at least one way did not depend significantly on situation, $D(1) = 0.11, p = .74$, or group, $D(3) = 4.46, p = .22$ (footbridge: 72%, switch: 70%).

To provide an indication of the difficulty of the standard situations, the tendency to reflect conflict was compared between the standard and the 1 scratched vs. 5 killed situations. Among the situations used in the present study, the 1 scratched vs. 5 killed situations were expected to be the most straightforward. As expected, participants were significantly less likely to express difficulty in the 1 scratched vs. 5 killed situations than in the standard situations, $D(1) = 94.67, p < .001$ (1 scratched vs. 5 killed footbridge: 45%, 1 scratched vs. 5 killed switch: 31%).

Summary: Conflict in judging standard situations. The findings revealed that participants found both the footbridge and the switch situations to involve conflicts. Most
expressed at least one form of difficulty, protesting consequences, considering alternative viewpoint, or accepting opposite decisions. This finding is especially noteworthy since participants were not asked explicitly whether they accepted the consequences of their evaluation, nor were they prompted to consider the alternative viewpoint when justifying their evaluation.

Justifications of evaluations about non-standard situations. Table 8 summarizes the justifications provided for evaluations about the 1 scratched vs. 5 killed, 1 killed vs. 5 family killed, 1 dog killed vs. 5 people killed, and 1 killed vs. 1 killed situations.

\[\text{\inserttable8}\]

1 scratched vs. 5 killed: Changing characteristics of the consequences of being pushed. Participants typically justified “alright” evaluations by value of life based on numbers or value of life based on rights justifications (80% - 85% of participants). These justifications were used significantly more often to justify “alright” evaluations than “not alright” evaluations, value of life based on numbers: alright: 82%, not alright: 15%, \(D(1) = 53.73, p < .001\), value of life based on rights: alright: 86%, not alright: 56%, \(D(1) = 12.41, p < .001\). The value of life based on rights justifications used to explain why intervention was not alright were typically statements about the intrinsic value of human life. For instance, one participant said it was alright to intervene in the 1 scratched vs. 5 killed switch situation because “you’re not killing [the person pushed in front of the train]: The person will survive” (male, 21 years).

The few participants who judged that it was not permissible to intervene in the 1 scratched vs. 5 killed situation primarily used value of life based on rights justifications (56%,
typically statements about lacking the right to force someone else to sacrifice their health), responsibility for consequences of actions (33%, vs 1% for “alright”), evaluation effect: $D(1) = 37.28, \ p < .001$, victim involvement (18%, vs. 1% for “alright”), $D(1) = 15.32, \ p < .001$, or consequences to self (28%, vs. 3% for “alright”), $D(1) = 19.34, \ p < .001$. Responsibility for consequences of actions justifications were also more common in the footbridge situation (9%) than in the switch situation (1%), $D(1) = 6.64, \ p = .001$. There were no other significant effects of evaluation, situation, or group, $ps > .05$.

1 killed vs. 5 family killed: Changing characteristics of the people saved. As shown by Table 8, evaluations that it was permissible to intervene in the 1 vs. 5 family situation were most commonly justified by consequences to self (e.g. references to particular obligations and relationships to one’s family members), evaluation effect (alright: 76%, not alright: 25%): $D(1) = 54.74, \ p < .001$, and value of life based on numbers justifications (alright: 53%, not alright: 13%), evaluation effect: $D(1) = 41.97, \ p < .001$. In the footbridge situation, one person justified her “alright” evaluation in the following way:

“One situation is just four children dying, your sister dying. And the other one is just killing someone. So, the only person who’s going to feel really guilty is you. So, I think your guilt versus saving five people. I think it’s worth more to risk your life and your emotional wellbeing for the people you love” (female, 23 years).

In contrast, the other justification categories were more commonly used as reasons as to why intervention was not permissible: value of life based on rights (alright: 6%, not alright: 70%), evaluation effect: $D(1) = 99.30, \ p < .001$, victim involvement (alright: 10%, not alright: 23%), $D(1) = 9.27, \ p = .002$, natural course of events (alright: 3%, not alright: 26%), $D(1) = 32.98, \ p < .001$, and responsibility for consequences of actions (alright: 2%, not alright: 42%),
There were also significant effects of situation on the use of *natural course of events*, \( D(1) = 4.56, p = .03\), and *consequences to self*, \( D(1) = 6.74, p = .009\), both of which were overall more common in the footbridge than switch situation. There were also significant group differences for two justification types: *Consequences to self* justifications, \( D(3) = 14.22, p = .003\), were more common among adolescents (79%) than among college student participants online (63%), \( W(1) = 4.27, p = .039\), and adults (42%), \( W(1) = 11.02, p = .001\). These justifications were also more common among college students interviewed in person (63%) than among adult participants, \( W(1) = 4.52, p = .033\). There were no other significant group differences for the use of *consequences to self* justifications, \( Ws(1) < 3.12, ps > .07\).

References to *victim involvement* also dependent significantly on group, \( D(3) = 8.83, p = .032\). These justifications were more common among adolescent participants (19%) than among undergraduates interviewed online (5%), \( W(1) = 4.60, p = .032\). Unexpectedly, these justifications were also more common among adults (21%) than among students interviewed online, \( W(1) = 6.31, p = .012\). The difference between college students interviewed in person (17%) and students interviewed online was not significant, \( W(1) = 3.37, p = .066\), and there were no other significant differences between groups, \( Ws(1) < 0.11, ps > .73\). There were no other significant effects of situation or group on other justification types, \( ps > .05\).

**1 Dog killed vs. 5 people killed: The role of the value of human life.** Participants who judged it as alright to sacrifice the dog to save the five workers typically used *value of life based on numbers* justifications (95% in footbridge and switch, vs. 8% - 5% for “not alright”), evaluation effect: \( D(1) = 127.24, p < .001\). Participants often said the human lives were of more worth than the life of a dog not just because of the number of lives directly saved but also because of the broader consequences of a human life (similarly in line with utilitarian thinking).
For instance, in explaining why it was permissible to push the dog off the bridge to save the workers, one participant said: “[F]ive human lives are more valuable than the dog. […] [They] have more to contribute to the world than one dog” (male, 25 years). In contrast, the minority of participants, who judged that intervention was not permissible typically referred to value of life based on rights, evaluation effect: $D(1) = 35.98, p < .001$, victim involvement, $D(1) = 54.42, p < .001$, natural course of events, $D(1) = 14.49, p < .001$, or responsibility for consequences of actions, $D(1) = 66.25, p < .001$. There were no other significant effects of evaluation, situation, or group, $ps > .05$.

1 killed vs. 1 killed: Distinctions between intervention and non-intervention when consequences are the same. Because of the low rate of “alright” evaluations, only justifications for “not alright” evaluations were analyzed. The most common justification types used to explain why intervention was not permissible was value of life based on rights justifications (58% for footbridge, 60% for switch) and natural course of events (52% in footbridge, 67% for switch). For instance, one participant explained why it was better not to intervene in the 1 killed vs. 1 killed switch situation as follows:

P: Because that’s the way it was meant to happen. That’s the way the accident was gonna occur and I shouldn’t interfere with that.

I: Why is it alright to let the one person die?

P: It’s not alright but it’s out of my control. If the only way I could save him would be by killing someone else then I wouldn’t want to be involved with that (female, 18 years).

As for the standard situations, the use of several justification types differed between footbridge and switch situations: Victim involvement, $D(1) = 19.19, p < .001$, and responsibility for consequences of actions, $D(1) = 6.62, p = .01$, were more common for the footbridge
situation, whereas natural course of events was more common for the switch situation, $D(1) = 9.88, p = .002$.

There were also significant effects of group for several justification types. Value of life based on numbers justifications, $D(3) = 22.97, p < .001$, were more common among adolescents (44%) than among students interviewed online (26%), $W(1) = 4.02, p = .045$, and adults (7%), $W(1) = 14.87, p < .001$. Value of life based on numbers justifications were also more common among students interviewed in person (33%) than among adults, $W(1) = 10.63, p = .001$, and among students interviewed online than among adults, $W(1) = 8.54, p = .003$. There were no other significant group differences for value of life based on numbers justifications in the 1 killed vs. 1 killed situations, $W_{s}(1) < 1.00, ps > .31$.

There was also a significant group effect for victim involvement, $D(3) = 43.21, p < .001$. These justifications were more common among adolescents (54%) than among college students interviewed online (8%), $W(1) = 29.31, p < .001$, and adults (21%), $W(1) = 14.36, p < .001$. Similarly, these justifications were more common among college students interviewed in person (44%) than among college students interviewed online, $W(1) = 20.36, p < .001$, and adults, $W(1) = 7.22, p = .007$. Finally, victim involvement justifications were more common among students interviewed online than among adults, $W(1) = 5.97, p = .015$. Adolescents and students interviewed in person did not differ significantly, $W(1) = 1.10, p = .29$.

Lastly, there was a significant effect of group on the use of consequences to self justifications, $D(3) = 12.70, p = .005$. These justifications were less common among students interviewed online (3%) than among adolescents (23%), $W(1) = 9.49, p = .002$, students interviewed in person (12%), $W(1) = 4.19, p = .041$, and adults (15%), $W(1) = 5.78, p = .016$. There were no other significant differences in the use of consequences to self justifications, $W(1)$. 
< 1.75, \( p_s > .18 \). There were no other significant effects of evaluation, situation, or group, \( p_s > .05 \).

*Summary: Justifications of evaluations about non-standard situations.* In sum, justifications for evaluations of non-standard situations were overall consistent with the patterns found in the standard situations. When participants said that it was not permissible to intervene, they were especially likely to reference the number of lives saved (*value of life based on numbers*). In contrast, when judging that intervention was not permissible, participants often reasoned about *value of life based on rights, victim involvement, natural course of events, responsibility for consequences, and consequences to self*.

On top of these general trends, and in line with our predictions, people’s reasoning was also sensitive to the unique features of each non-standard situation pair. For instance, in the 1 scratched vs. 5 killed situations, most participants considered the non-lethal consequences of pushing or flipping the switch, and referenced the intrinsic rights and value of the people being saved (*value of life based on rights*) when explaining why it was permissible to intervene. In the 1 vs. 5 family situations, people’s reasoning reflected their relationship to the individuals being saved: Most people referenced *consequences to self* when explaining why it was permissible to sacrifice a worker to save five family members. In response to the 1 dog killed vs. 5 people killed situations, the vast majority of participants judged intervention permissible because they thought five human lives were worth more than a dog’s life (*value of life based on numbers*). In contrast, the participants who judged intervention as wrong gave priority to moral concerns for the dog, particularly *value of life based on rights, victim involvement, natural course of events, and responsibility for consequences*. Finally, in the 1 killed vs. 1 killed situations, when the consequences in terms of number of lives lost were the same for intervention and non-
intervention, participants overwhelmingly gave priority to the negative moral ramifications of intervention. Participants primarily referenced *value of life based on rights, victim involvement* (especially in footbridge variant), *natural course of events* (especially in switch variant), or *responsibility for consequences* (especially in footbridge variant) when explaining why it would be wrong to sacrifice one worker to save another.

**Discussion**

Study 1 provided evidence that people reasoned about the standard switch and footbridge situations in qualitatively different ways. Most participants considered the standard switch situation primarily in utilitarian (*value of life based on numbers*) terms, and correspondingly judged intervention as permissible. In contrast, in response to the footbridge situation, participants tended to bring in considerations about *value of life based on rights, victim involvement, responsibility for consequences of action, and natural course of events*, often judging intervention as not permissible. Participants who judged the footbridge and switch situations differently explained the difference by reference to what they saw as unique and morally relevant features of each scenario. Importantly, however, a substantial number of participants judged interventions in both standard situations as non-permissible and provided similar justifications for both situations. The distinct considerations brought to bear on trolley car situations, including non-utilitarian concerns with the value of life, the roles and obligations assumed by those involved, and the integrity of actors, aligned with what several philosophers (e.g., Foot, 1967; Williams, 1973) have proposed.

As importantly, Study 1 demonstrated that the conflicts involved in violating the value of life to preserve that very value are salient to most people. Participants generally viewed the
standard trolley car situations as multifaceted and were willing to consider and even accept evaluations opposite of their own (consistent with Kahane, 2015; Sheskin & Baumard, 2016).

Responses to the non-standard situations demonstrated that participants were sensitive to variations in key elements of the trolley car situations. At all ages, participants’ reasoning and evaluations varied with situational features such as the consequences to the victim (1 scratched vs. 5 killed), specific responsibilities toward family members (1 killed vs. 5 family killed), the value of human life (1 dog killed vs. 5 people killed), and the added personal responsibilities taken on by intervening in the situation (1 killed vs. 1 killed).

Somewhat surprisingly, Study 1 revealed few age differences. Patterns of evaluations and reasoning were similar for adolescents, undergraduate students, and other adults. In fact, there were more differences between participants interviewed in person (adolescents and undergraduate samples) and participants interviewed via online surveys (undergraduate and other adult [mTurk] samples) than there were age-related differences. For instance, perhaps unsurprisingly, participants provided more reasons during in-person interviews than in online surveys. This finding, along with the completeness of the interview transcripts compared to the generally shortened responses to online surveys, highlights how in-person interviews can yield a more comprehensive portrayal of how individuals think about social situations than do online surveys. In addition, the group differences suggest that adolescents, sometimes at least, judge and reason about life and death in similar ways to adults.

The similarity of responses between the four groups is especially surprising since the groups also differed in ways other than age and interview method. The undergraduate samples were recruited from a research university in the Western United States, the adolescent sample came from a high school in New England, and the adults were recruited through Amazon
Mechanical Turk were distributed across the United States. The groups also differed somewhat in the distribution of gender and ethnicity, and we did not observe any statistically significant effects of these variables. Past studies have reported similar consistency in responses to trolley car situations across a variety of individual and group characteristics (e.g. Hauser et al., 2007). It will be informative to use the methods adopted here to study reasoning about sacrificing and saving lives in additional cultural groups. We expect that the moral considerations made by the present participants will be evident in most adult populations, but that the frequency and prioritization of each consideration may vary (Kohlberg, 1971; Turiel, 2002).

The findings from Study 1 were consistent with our hypotheses about the centrality of reasoning and challenge dual-process and intuitionist accounts of responses to trolley car situations (Greene, 2014). The study documented that both adolescents and adults reason about multiple moral considerations when evaluating trolley car situations, not just the utilitarian principle of maximizing the number of lives saved (Greene, 2014). The evidence from Study 1 that people were conflicted about both switch and footbridge situations also contrasts with proposals by dual-process explanations of responses to trolley car situations. Typically, dual-process accounts imply that people experience little conflict when judging that it is permissible to intervene in the switch situation, and automatically think that it is wrong to intervene in the footbridge situation.

Study 1 investigated several aspects of participants’ responses to trolley car situations. Each of the reasoning categories used merit further investigation. One type of consideration that has rarely been discussed in the empirical literature, and that participants referenced more frequently than we had expected, concerned the victim’s involvement in the situation.
Victim involvement justifications mainly came in two forms. Some participants referenced the innocence of the person to be sacrificed, for instance noting that the person on the footbridge was entirely outside the dangerous situation. Other times, participants referenced the accepted risk of the persons to be sacrificed, typically noting that the railway workers had chosen to work on the railway. Victim involvement justifications are noteworthy not only because they have received little attention in the empirical literature on trolley car situations, but also because they illustrate how participants may need to fill in background assumptions about hypothetical situations to form evaluations (Elster, 2011). Insofar as the situation descriptions do not explicitly state how potential victims were involved in the situation, and victim involvement is seen as relevant to evaluations, participants must implicitly or explicitly construct assumptions. Whereas researchers sometimes discount such added assumptions in coding participant justifications (Hauser et al., 2007), it is not clear that participants could evaluate hypothetical situations without making at least some background assumptions not given in the situation descriptions: Descriptions of situations are usually short and yet innumerable considerations could potentially be relevant for the evaluation of the actions.

To further investigate the role of victim involvement considerations, we conducted a second online interview study with undergraduates in which we manipulated the role of the potential victims in the trolley car situations. In addition, Study 2 also served as an experimental test of our proposal that the reasons provided by participants informed their evaluations. If participants said victim involvement mattered, yet did modify their evaluations in response to changes in victim involvement, this would support the proposal that these justifications were mere “post hoc rationalizations” that did not influence participants’ evaluations, as proposed by intuitionist and dual process accounts (Greene, 2014; Haidt & Bjorklund, 2008a). In contrast, if
participants’ evaluations were affected by victim involvement, this would support our proposal that the reasons provided by participants did indeed inform their evaluations.
Chapter 3: Study 2: Varying the Involvement of the Potential Victims

In Study 1, many participants made references to victim involvement in explaining why it was not permissible to intervene in the footbridge situation. Victim involvement justifications highlighted a key difference between the way the standard footbridge and switch situations were presented in their original form. In the switch situation the person sacrificed was described as a railway worker, whereas in the footbridge situation the person was not described as a railway worker. Therefore, participants held that the persons on the main or side track had accepted a position of responsibility or risk, whereas the person on the footbridge had not. Participants consequently thought it was wrong to bring someone uninvolved in the situation into a position of danger, for instance in the standard footbridge situation.

Past philosophical treatments of trolley car situations have proposed that victim involvement is a relevant consideration for evaluations of trolley car situations (Foot, 1967; Thomson, 1985). We previously quoted Thomson (1985)’s hypothetical example of five mafia members on the main track who had tied a person to the side track in order to kill him. This, according to Thomson, would provide a reason for not flipping the switch, but instead letting the train hit the five mafia members. Some empirical research has also indicated that people’s evaluations are affected by characteristics of the persons to be sacrificed (Petrinovich, O’Neill, & Jorgensen, 1993).

Study 2 investigated how changes in victim involvement affect people’s evaluations and justifications about trolley car situations. In one variant (culprit killed vs. 5 workers killed), the person to be sacrificed to save the five workers had deliberately caused the train to head down the track in order to kill the five workers. Thus, the single person on the bridge or sidetrack was far more involved and culpable than the five workers: The single person had caused the
dangerous event whereas the workers had at most accepted the risk of being on a track. The inclusion of the culprit killed vs. 5 workers killed situation also allowed for a further test of people’s aversion to pushing. Study 1 showed that participants were more likely to judge pushing as permissible when they were pushing a person to his death to save family members, or when pushing a dog, or when the person pushed was merely scratched or bruised. In the culprit killed vs. 5 workers killed footbridge situation of Study 2, the physical consequences of pushing were the same as in the standard footbridge (one person killed, as opposed to a dog killed or a person scratched or bruised) and there was no personal relationship (e.g. to family members) involved.

We also included two situation pairs in which the role of the person who would be sacrificed would be similar in the switch and footbridge variants. In one variant (worker killed vs. 5 workers killed) the person to be sacrificed was always a worker: for the footbridge situation it was stated that the person to be pushed was “a large railway worker,” while the switch situation was the same as in the standard version. In another variant (non-worker killed vs. 5 workers killed) the person to be sacrificed was always outside the situation (i.e., clearly a non-worker): the description in the footbridge situation was the same as in the standard situations, while for the switch situation the person was described as laying in a hammock away from the tracks.

Finally, Study 2 differed from Study 1 in using only one situation pair per participant. This ruled out the possibility that participants could notice differences between the situation pairs and let such differences influence their evaluations and justifications.

Study 2 used online interviews with undergraduate participants from the same research participant pool as the undergraduate student participants in Study 1 (although no participant
participated in both studies). Thus, Study 2 was most directly comparable to the online interviews with undergraduate students in Study 1.

**Method**

**Participants.** Study 2 had 144 participants ($M_{age} = 20.3$ years, $SD_{age} = 2.6$ years, 66% female). In addition, data from three participants was collected but not analyzed because they did not respond to all the interview questions. Participants were recruited from a research participation pool from the same large public university in the Western United States from which the undergraduate samples in Study 1 were recruited. The race and ethnicity distribution of the sample was: 46 percent Asian, 32 percent White non-Hispanic, 13 percent Hispanic, and 9 percent other. Participants received course credit for their participation.

**Materials and procedures.** Each participant participated in an online survey with the same format, and using the same questions, as in Study 1. Participants were randomly assigned to receive one of the three situation pairs (culprit killed vs. 5 workers killed, worker killed vs. 5 workers killed, or non-worker killed vs. 5 workers killed: 48 participants received each situation pair). The presentation of switch and footbridge variants was counterbalanced. Unlike Study 1, Study 2 did not ask participants to compare their evaluations about footbridge and switch situations.

Situations were described as in Study 1, with the following exceptions: In the non-worker killed vs. 5 workers killed situation, the last part of the switch variant read: “The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate, unfinished track and continue past the end of that track, down a hill, and into a yard where a man is sleeping in a hammock. The man in the hammock would be killed if you flip the switch.” In the worker killed vs. 5 workers killed situation, the last part of the “footbridge” variant read: “In
front of you, right next to the track, there is a large railway workman. The only way to save the five workmen is to push the single workman, who would fall onto the tracks and stop the train. The single workman will be killed if you push him.” In the culprit killed vs. 5 workers killed situation, the person on the sidetrack (switch variant) or on the footbridge (footbridge variant) was described as the man who “deliberately caused the train car to head down the track because he wanted to kill the five workers.” The complete texts of the situation descriptions were as follows:

**Culprit killed vs. 5 workers killed**

*Switch variant.* You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a switch. The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate track where there is a man. That man on the side track actually deliberately caused the train car to head down the track because he wanted to kill the five workers. The man will be killed if you flip the switch.

*Footbridge variant.* You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a large stranger. The only way to save the five workmen is to push the stranger, who would fall onto the tracks and stop the train. That stranger on the bridge actually deliberately caused the train car to head down the track because he wanted to kill the five workers. The stranger will be killed if you push him.
Moral reasoning 80

**Worker killed vs. 5 workers killed**

*Switch variant (same as standard switch).* You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a switch. The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate track where there is a single railway workman. The single workman will be killed if you flip the switch.

*“Footbridge” variant.* You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. In front of you, right next to the track, there is a large railway workman. The only way to save the five workmen is to push the single workman, who would fall onto the tracks and stop the train. The single workman will be killed if you push him.

**Non-worker killed vs. 5 workers killed**

*Switch variant.* You are standing near a train track as you see a runaway train car moving down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a switch. The only way to save the five workmen is to flip the switch, which would divert the train onto an alternate, unfinished track and continue past the end of that track, down a hill, and into a yard where a man is sleeping in a hammock. The man in the hammock would be killed if you flip the switch.

*Footbridge variant (same as standard footbridge).* You are standing on a pedestrian walkway that arches over a train track as you see a runaway train car moving
down the track. In the path of the train are five railway workmen, who will surely be killed unless you do something. Next to you there is a large stranger. The only way to save the five workmen is to push the stranger, who would fall onto the tracks and stop the train. The stranger will be killed if you push him.

Coding and data analysis. Coding and data analysis were as in Study 1, except that the Generalized Linear Mixed Models (GLMMs) did not include fixed effects of group (since all participants in Study 2 were undergraduates participating in an online survey). The culprit killed vs. 5 workers killed situations often elicited references to how the person on the footbridge or sidetrack was responsible for the situation (because he had caused the train to head down the track in an attempt to kill the five workers). Although such justifications were unique to the culprit scenario, they were classified as victim involvement justifications. Like statements about accepted risk of railway workers, references to the responsibility of the culprit indicated that the person’s death was acceptable because he had accepted the possibility of the harmful events. Twenty percent of interviews were randomly selected for double-coding in order to assess intercoder agreement. Agreement for evaluations, as measured by Cohen’s κ, was κ = .98. Average agreement for justifications was κ = .80 (range: .69 to .95).

Results

Evaluations. There was a significant effect of situation variant on whether people judged intervention as permissible for the footbridge, $D(2) = 42.71, p < .001$, and switch situations, $D(2) = 20.87, p < .001$. For footbridge situations, the majority (73%) judged it permissible to intervene in the culprit killed vs. 5 workers killed situation. In contrast, few judged it permissible to intervene in the worker killed vs. 5 workers killed (17%) and non-worker killed vs. 5 workers killed (i.e., standard; 19%) situations. For the switch situations, 92% judged it as permissible to
intervene in the culprit killed vs. 5 workers killed situation, 65% judged intervention permissible in the worker killed vs. 5 workers killed (standard) situation, and 52% judged intervention permissible in the non-worker killed vs. 5 workers killed situation (Table 9).

As can be seen in Table 9, participants were overall more likely to say it was permissible to intervene in the switch situations (69%) than in the footbridge situations (34%), \( D(1) = 80.64, p < .001 \). This difference was significant for all three situation types, \( D_s > 22, p_s < .001 \), and the interaction between footbridge vs. switch and situation type was not significant, \( D(2) = 4.02, p = .13 \).

**Justifications provided in response to each situation.** As in Study 1, the justifications provided in Study 2 varied as a function of evaluation and situation (Table 10). Below, we first summarize the evaluations and most common justifications provided for each situation type. We then present the statistical analyses of how justification categories differed between situations.

**Culprit killed vs. 5 workers killed.** A main finding of Study 2 was that the majority of participants judged intervention in the culprit killed vs. 5 workers killed situation to be permissible – even though it involved pushing another person. Most judged flipping the switch (92%) or pushing the person from the footbridge (73%) to be permissible when the person to be sacrificed had deliberately caused the train to head toward the people on the track.
A large proportion of participants used *victim involvement* justifications to explain why it was permissible to sacrifice the culprit to save the five workers (88% in footbridge, 78% in switch). One common type of *victim involvement* justification was references to how the culprit had sought to bring about these dangerous circumstances. (As noted above, these were considered *victim involvement* justifications since they referred to the victim’s implicit or explicit acceptance of the harmful circumstances.) One participant said it was alright to push the culprit off the footbridge to save the workers because “he's considered a potential killer, planning to kill those five men. In a way, it would be a form of self-defense to save the five men” (male, 16 years).

Many participants also used *value of life based on numbers of lives* justifications to explain why it was permissible to sacrifice the one person to save the five (64% in footbridge, 67% in switch). Among the limited number of participants who said intervention was not alright, most referred to either *value of life based on rights* (40% in footbridge, 67% in switch), *responsibility for consequences of actions* (47% in footbridge, 0% in switch), or *consequences to self* (27% in footbridge, 33% in switch).

**Worker killed vs. 5 workers killed.** In the worker killed vs. 5 workers killed switch situation, 65 percent said intervention was alright, whereas in the worker killed vs. 5 workers killed footbridge situation, 19 percent said intervention was permissible. An explanation is not evident as to why the percentage in the worker killed vs. 5 workers killed switch situation was somewhat higher than the comparable percentage for online undergraduate participants in Study 1 (55%). We note, however, that the Study 2 percentage is closer to that found in previous studies (Cushman et al., 2006; Hauser et al., 2007). Most importantly, participants were far less
prone to say that intervention was permissible for the worker killed vs. 5 workers killed situations than for the culprit killed vs. 5 workers killed situations.

As in Study 1, the most common reason provided for why intervention is permissible was *value of life based on numbers* justifications (100% in footbridge and switch). When justifying why intervention was not permissible, participants predominantly referenced *value of life based on rights* (61% in footbridge, 65% in switch), *natural course of events* (34% in footbridge, 35% in switch), and *responsibility for consequences of actions* (68% in footbridge, 53% in switch). In contrast, participants rarely used *victim involvement* to explain why it was not alright to intervene in the worker killed vs. 5 workers killed situations (13% in footbridge, 12% in switch).

**Non-worker killed vs. 5 workers killed.** In the non-worker killed vs. 5 workers killed switch situation, 52 percent said intervention was permissible, compared to 65 percent in the worker killed vs. 5 workers killed switch situation and 92 percent in the culprit killed vs. 5 workers killed switch situation. For the footbridge variants, percentages for non-worker killed vs. 5 workers killed (17%) and worker killed vs. 5 workers killed (19%) variants were highly similar and appreciably lower than the culprit killed vs. 5 workers killed footbridge situation (73%).

The most common justification for why intervention was permissible in the non-worker killed vs. 5 workers killed situations was *value of life based on numbers* (100% in footbridge, 88% in switch). When reasoning as to why intervention was not alright in the non-worker killed vs. 5 workers killed situation, participants referenced *value of life based on rights* (60% in footbridge, 62% in switch), *responsibility for consequences of actions* (52% in footbridge, 38% in switch), *natural course of events* (30% in footbridge, 33% in switch), and, of particular interest here, *victim involvement* (20% in footbridge, 25% in switch).
Statistical analyses of each justification category. In this section, we present the statistical analyses of situation and evaluation effects on the use of each justification category (see Table 10).

Value of life based on numbers. For both the switch and footbridge situations, value of life based on numbers justifications were more common when participants judged intervention as permissible, $D(1) = 172.71.67, p < .001$. On average, 80% used value of life based on numbers justifications after giving “alright” evaluations, whereas only 3% did so when they gave evaluations of “not alright”. There was also a significant effect of situation, $D(2) = 15.52, p < .001$. When judging situations as permissible, participants used value of life based on numbers justifications in 100% of worker killed vs. 5 workers killed situations, 88% - 100% of non-worker killed vs. 5 workers killed situations, and 64% - 67% of culprit killed vs. 5 workers killed situations. The effect of footbridge vs. switch was not significant, $D(1) = 0.11, p = .73$, when controlling for evaluation and situation type.

Value of life based on numbers justifications appeared to be used less in the culprit killed vs. 5 workers killed scenario because participants focused on the responsibilities of the culprit. For instance, the following participant explicitly disregarded the number of lives involved:

I: Why is it alright to let the one person [culprit] die?

P: It’s less about the number of people than the type of person that gets killed. A person in society that decides to kill others has broken the central part of the social contract and thus sacrifices the liberties that keep him alive (female, 21 years).

Value of life based on rights. Value of life based on rights justifications were the most common when participants said intervention was not permissible, $D(1) = 73.87, p < .001$ (used
for 59% of “not alright” evaluations and 4% of “alright” evaluations). There were no significant effects of switch vs. footbridge, $D(1) = 1.21$, $p = .27$, or situation type, $D(2) = 0.26$, $p = .88$.

**Victim involvement.** Victim involvement justifications depended significantly on situation type, $D(2) = 56.61$, $p < .001$. In the culprit situations, the majority of participants who said it was permissible to intervene used victim involvement justifications (88% in footbridge, 78% in switch). As noted above, these participants typically referred to how the culprit had tried to kill others and that it is therefore permissible to push him in front of the train to save the others. In contrast, for the non-worker killed vs. 5 workers killed and worker killed vs. 5 workers killed situations, victim involvement justifications were almost exclusively used to justify why it was not permissible to intervene. Such justifications were more common among participants saying it was not permissible to intervene in the non-worker killed vs. 5 workers killed situations (20% in footbridge, 25% in switch) than among participants saying it was not permissible to intervene in the worker killed vs. 5 workers killed situations (13% in footbridge, 12% in switch). In the non-worker killed vs. 5 workers killed situations, participants were thus more likely to note that the person to be sacrificed, unlike the five workers, was not involved in the situation. Explaining why it was not alright to flip the switch and redirect the train from the five workers to the person in the hammock, one participant stated: “As workers in a hazardous environment, [the five workers] should be wary of such occurrences. Meanwhile, the innocent man sleeping elsewhere has nothing to do with the train, and should not be involved” (male, 19 years). There were no significant effects of evaluation, $D(1) = 0.00$, $p = .95$, or switch vs. footbridge, $D(1) = 1.45$, $p = .23$, after controlling for situation type. In sum, the use of victim involvement justifications suggested that, as they were judging the situation, people did indeed perceive, and consider, the involvement of the agents in bringing about the potentially harmful situation.
**Natural course of events.** Natural course of events justifications were more common for evaluations of not alright (31%) than for evaluations of alright (7%), $D(1) = 14.15, p < .001$. When controlling for evaluations, there were no significant effects of switch vs. footbridge, $D(1) = 0.001, p = .98$, or situation type, $D(2) = 5.67, p = .06$.

**Responsibility for consequences of actions.** Participants were more likely to refer to responsibility for consequences of actions when they said intervention was not alright (53%) than when they said intervention was alright (6%), $D(1) = 45.97, p < .001$. There was also a significant effect of situation type, $D(1) = 4.54, p = .033$, as participants were overall more likely to use such justifications when reasoning about footbridge situations (41%) than about switch situations (16%). There was not a significant effect of situation type, $D(2) = 4.73, p = .09$.

**Consequences to self.** Finally, participants were more likely to refer to consequences to self when justifying why intervention was not alright (20%) than when justifying why intervention was alright (5%), $D(1) = 10.11, p = .001$. There were no significant effects of switch vs. footbridge, $D(1) = 0.02, p = .89$, or situation type, $D(2) = 0.16, p = .92$.

**Discussion**

Study 2 provided further evidence that participants’ responses to trolley car situations are informed by considerations about the involvement of potential victims. Participants were far more accepting of intervention when the person on the sidetrack or the footbridge had set off the trolley to kill the five workers (*culprit killed vs. 5 workers killed*) than in the other situations. In the switch situations, there was also a tendency for participants to be more likely to deem intervention as permissible when the person to be sacrificed was already on the track than when the person was in a hammock out of sight of the track. These considerations were reflected in people’s justifications. Participants were more likely to use victim involvement justifications
when the person to be sacrificed and persons to be saved differed in levels of involvement (culprit killed vs. 5 workers killed, non-worker killed vs. 5 workers killed) than when they had similar levels of involvement (worker killed vs. 5 workers killed). When potential victims were all workers and had similar levels of involvement, people were particularly prone to reason in terms of the number of people who would live and die.

The participants and methods of Study 2 were most similar to those of the undergraduate sample participating in online interviews in Study 1. The two groups were drawn from the same participant pool at a research university and were similar in distribution of age, gender, and ethnicity. Further suggesting that the two samples were similar, responses to the standard switch situation (worker vs. 5 workers in Study 2) and the standard footbridge situation (non-worker vs. 5 in Study 2) were highly similar.

The findings of Study 2, like those of Study 1, run counter to past claims that people mainly reason about trolley car situations in utilitarian terms (Greene, 2014). Participants justified their evaluations by references to several non-utilitarian considerations, such as the involvement of the potential victims. Moreover, when we experimentally manipulated the involvement of potential victims in Study 2, participants’ evaluations and justifications shifted accordingly, indicating that their responses were indeed informed by victim involvement. Perhaps most strikingly, almost three out of four participants thought it was permissible to push the culprit off the footbridge, predominantly justifying this evaluation with references to the culprit’s involvement in bringing about the dangerous situation.

The experimental design of Study 2 makes it unlikely that participants’ justifications were mere post hoc rationalizations (Greene, 2014; Haidt & Bjorklund, 2008a). The idea of post hoc rationalizations is that people would be creating justifications to fit their evaluations, and
their justifications should therefore be disconnected from their pattern of evaluations. By contrast, we found a high degree of consistency between evaluations, justifications, and situational features. In Study 1, participants stated that their evaluations were informed, in part, by the involvement of victims in the dangerous situation. By experimentally manipulating the involvement of potential victims, Study 2 showed that participants’ evaluations were indeed influenced by such involvement. This consistency between evaluations and justifications indicates that participants were aware of and endorsed the concerns on which their evaluations were based, providing evidence for moral reasoning (as defined above) rather than post hoc rationalizations.

Although manipulations of victim involvement significantly influenced evaluations and justifications, it did not remove a difference between the footbridge and switch situations. In the culprit killed vs. 5 workers killed situations, in which the majority of participants stated that intervention was permissible in both the footbridge (73%) and switch variants (92%), participants were still significantly more likely to accept intervention in the switch variant. Does this mean that concerns with victim involvement do not explain the differences in people’s responses to footbridge and switch variants of the trolley car situations? Yes and no. On the one hand, participants’ responses indicated that they saw a variety of differences between the switch and footbridge situations, and that victim involvement was only one such difference (others included concerns with rights, responsibilities, and personal integrity). No one of these differences by itself will fully account for why people are more likely to view intervention as permissible for switch variants than for footbridge variants (Greene et al., 2009). At the same time, Study 2 indicated that participants’ evaluations were responsive to changes in victim
involvement, and that concerns with such involvement is one reason why people treat the
footbridge and switch situations differently.
Chapter 4: General Discussion

The observed patterns of evaluations, justifications, and conflicts in response to trolley car situations provide, in our view, convincing evidence that adolescents and adults reasoned in moral terms about sacrificing and saving lives. Their reasoning involved several considerations, not only the number of lives lost and saved, but also the value of life based on rights, victim involvement, and the role and responsibilities of the potential intervener. In reasoning about sacrificing and saving lives, participants often considered multiple and sometimes conflicting considerations.

These research findings bridge two literatures that, most of the time, have been kept separate. First, the present studies included new dimensions in research on the development of moral evaluations and orientations. Second, the findings address debates in moral psychology, proposing an account in which moral reasoning is central to how people form moral evaluations and challenging current dual-process and intuitionist views.

In this concluding chapter, we begin by briefly reviewing the psychological and philosophical background for the research. We then recapitulate the main findings in relation to the three main questions posed in the Introduction: (1) How do adolescents and adults reason about trolley car situations? (2) Are people conflicted about their evaluations of trolley car situations? (3) Are people’s evaluations and justifications sensitive to morally relevant variations of the situations? Next, we discuss implications of the findings for research and theories about the development of morality. Finally, we relate the findings to contemporary debates in moral psychology about relations among reasoning, intuitions, emotions, and actions.
Motivation and Main Conclusions

The present research was based on developmental research on how children and adults reason about social and moral decisions. A major finding of that research is that children and adults often deal with social situations that require the weighing and balancing of different moral and non-moral considerations (Killen & Smetana, 2015; Turiel, 2015). The studies presented in this Monograph were guided by long-standing propositions that individuals strongly hold conceptual and emotional concerns about the value of life (Colby & Kohlberg, 1987a, 1987b; Dworkin, 1993; Turiel & Killen, 2010).

Little was known from past research about how individuals develop moral concerns with the value of life, nor how they develop the ability to deal with conflicts involving sacrificing and saving lives. In the Introduction we described limitations in the work of Kohlberg (1963) on situations pitting the value of life against other considerations. It was clear that life and death were salient issues to adolescents (Balk & Corr, 2009; Rosengren et al., 2014; Slaughter, 2005), but we did not know how adolescents would reason about sacrificing lives to save others. To investigate reasoning about the value of life in situations involving sacrificing and saving lives among adolescents and adults, we interviewed participants about trolley car situations taken from the philosophical and psychological literatures (Edmonds, 2013; Foot, 1967; Greene, 2014; Thomson, 1985).

Even though the trolley car situations clearly bear on issues of life and death, prior research had not attended to those issues in a detailed way. Instead, negative reactions to directly killing one person to save five others were treated as a result of emotional and automatic reactions rather than understandings of moral principles. By using a broad range of situation
types and by systematically probing people’s responses, the present studies revealed information about how people think about trolley car situations and similar situations of conflict.

The situations used in this research entailed intense conflicts – situations that philosophers, such as Walzer (2007), refer to as “hard cases.” Walzer did not consider evaluations about the trolley car situations, but other types, such as evaluations about the use of torture as a means of saving lives. He discussed, as an example, the question of whether it would be permissible to torture someone such as a known terrorist in order to obtain information that would save the lives of many from a terrorist attack. As he put it (Walzer, 2007, p. 302), “[o]ne of my examples was the ‘ticking bomb’ case, where a captured terrorist knows, but refuses to reveal, the location of a bomb that is timed to go off soon in a school building.” Walzer regards this as a “hard case” requiring exceptions to the application of rules because it involves violating a moral goal in order to promote a moral goal. The question of whether it would be permissible to torture someone as a means of saving lives is often debated, with people on each side of the issue providing utilitarian and non-utilitarian reasons for their positions. Those maintaining that torture is wrong often do so on grounds such as preserving principles of human dignity, preserving rights, and preventing people in authority from violating agreed upon rules. Their non-utilitarian positions are not without reasons.

So what have we found? And what do the findings tell us about how individuals evaluate and reason about hard cases?

(1) *How do adolescents and adults reason about trolley car situations?* First, we found that reasoning about trolley car situations is not of just one kind: Evaluations and justifications provided in both studies indicated that participants reasoned about several features of the footbridge and switch situations, and responses were not simply based on differences between
pushing another and flipping a switch. In some cases, it was judged permissible to take a life to save five: when a switch is thrown that would redirect the train (standard, 1 killed vs. 5 killed); when the switch is thrown or a person is pushed that would save the lives of five family members; and when throwing the switch or pushing someone who was not an innocent person (i.e., a culprit who had intentionally caused the train to hit the five). Additionally, it was found that most judged it permissible to push another who would be injured but not killed (1 scratched vs. 5 killed). In other situations, most judged that it would be wrong to push another or throw the switch when sacrificing one life would save one life.

Clearly, the acts of throwing the switch or pushing another, as well as the utilitarian calculation of the benefit of saving more lives, are not the only features involved in arriving at decisions in these trolley car situations. It appears that adolescents, young adults, and older adults do take into account the saving of lives, but also consider responsibilities for whose lives are involved (as evidenced in the culprit and family situations). The focus on human lives seems evident in the evaluations that it is permissible to push a dog that would be killed, as well as in decisions that it is wrong to sacrifice one life to save one life.

Justifications in the various situations indicate that – in addition to considerations about the number of lives – non-utilitarian considerations are of substantial importance. These other considerations include the intentions of the actor, the responsibilities of those who would or would not be killed, the rights of individuals involved, the conflict between the principle that it is wrong to take actions that would kill another, and the general obligation to save lives. That these types of considerations were at work is demonstrated by the third salient feature of the findings: namely, that participants at all ages provided reasons for their evaluations, reasons which systematically aligned with evaluations in the different types of situations. Perhaps not
surprisingly, participants who judged that it was permissible to sacrifice one person to ensure the survival of five others typically referred to the number of lives saved (*value of life based on numbers*). By contrast, participants who judged that it was *not* permissible to sacrifice one person to save five others typically considered several relevant considerations, including the intrinsic value of life and rights of individuals (*value of life based on rights*), the victims’ involvement in the situation (*victim involvement*), the natural course of events, their responsibility for the consequences of intervening, and the consequences to themselves of intervening.

(2) Are people conflicted about their evaluations of trolley car situations? People’s responses to the standard (1 killed vs. 5 killed) trolley situations indicated that participants indeed found these situations to be hard cases. Their difficulty in evaluating the standard situations was evident in their consideration of reasons for the opposite evaluation, their refusal to accept the consequences of their chosen action, and their acceptance of others coming to a different evaluation about the situation. Participants were more likely to show such signs of difficulty when judging the standard situations than when judging the seemingly easier 1 scratched vs. 5 killed situations. In our interpretation, the presence of conflict is a consequence of the fact that people view trolley car situations as relating to multiple and sometimes conflicting considerations, not just a counting of the number of lives sacrificed and saved.

(3) Are people’s evaluations and justifications sensitive to morally relevant variations of the situations? The evaluations and justifications across Studies 1 and 2 indicated that people’s reasoning was attuned to relevant features of the situations. More people judged intervention as permissible in the standard (1 vs. 5) switch situation than in the standard footbridge situation. These two situations were perceived to differ along several dimensions deemed relevant to the evaluation of the acts. When asked why they judged the standard switch and footbridge
situations differently, participants pointed to several differences between the two situations that they deemed morally relevant. Participants noted that the person on the footbridge was “innocent” or outside the situation whereas the worker on the sidetrack was not (victim involvement), that pushing the person off the footbridge, but not flipping the switch, entailed taking responsibility for the person’s death (responsibility for consequences), and (possibly for the latter reason) anticipated more guilt or social repercussions if they pushed the person off the bridge than if they flipped the switch (consequences to self, see Table 5).

Responses to non-standard situations in both studies also revealed a close correspondence between situational features, justifications, and evaluations. For instance, when the individual to be pushed off the footbridge to save the five workers would be scratched and bruised but not killed (1 scratched vs. 5 killed) or was a dog (1 dog killed vs. 5 people killed), the majority of participants said it would be permissible to intervene even in the footbridge scenario (Study 1). Participants justified these evaluations by emphasizing not only the number of human lives saved but also, in 1 scratched vs. 5 killed, the intrinsic value of life as opposed to minor physical harm. In other words, when pushing the individual off the bridge infringed less on the individual’s intrinsic value and rights (because of the limited physical harm to a person or the perceived limited rights of a dog), the value of the lives and rights of the persons on the track took precedence.

Study 2 further demonstrated the systematic relations among situational features, justifications and evaluations. Importantly, Study 2 participants only responded to one situation pair, reducing the likelihood that they adjusted their evaluations and justifications to meet perceived experimenter expectations. When the person to be sacrificed was one who had intentionally attempted to kill the five workers (culprit killed vs. 5 workers killed), the vast
majority of participants judged it permissible to intervene and save the five others. There was a clear recognition that a culpable person does not have the same rights or merit the same considerations as “innocent” parties. Participants’ justifications for these evaluations were especially likely to refer to the culprit’s involvement in the situation (*victim involvement*). There were also relations between evaluations and justifications in the contrast between the non-worker killed vs. 5 workers killed and worker killed vs. 5 workers killed variants. When the person to be sacrificed was in a hammock away from the tracks (non-worker killed vs. 5 workers killed switch), as opposed to on the sidetrack (worker killed vs. 5 workers killed switch), participants were more likely to say that it was not permissible to flip the switch and to justify this evaluation by reference to the situational involvement of the person to be sacrificed (*victim involvement*). Evaluations for the footbridge situations with workers and non-workers did not vary and were associated mainly with reasons of intrinsic rights.

The correspondence between evaluations, justifications, and situational features supports the proposal that people’s justifications typically reflect considerations that inform people’s evaluations. If people said that they evaluated situations based on victim involvement, yet their evaluations did not change when victim involvement changed, such discrepancy would suggest that people were not truly concerned with victim involvement. In contrast, the pattern of evaluations and justifications across the different situations, most strikingly in Study 2, indicate that concerns with victim involvement did guide participants’ evaluations.

**The Development of Moral Concerns with Human Lives**

The pattern of findings summarized above characterized both adolescent and adult responses in Study 1. We found only minor differences between participant groups, and these differences were typically associated with methodological differences (online vs. in-person
interviews) rather than age differences between participant groups. These findings indicate that reasoning about multiple moral concerns regarding the value of life is generally present by adolescence.

The lack of developmental differences may seem surprising. However, we entered this research treating it as an open question whether we would see differences between adolescents and adults in responses to trolley car situations. Past research has indicated that, by adolescence, individuals incorporate multiple and sometimes conflicting considerations about rights, welfare, and fairness in multifaceted situations (Damon, 1975; Elenbaas et al., 2016; Helwig, 1995; Rizzo & Killen, 2016; Turiel & Nucci, 2017). The present research indicates that adolescents are not only sensitive to the same kinds of moral considerations about life as adults, including the number of lives involved, rights to life, victim involvement, and the roles and responsibilities of the intervener, but also appeared to coordinate these considerations to arrive at evaluations that were very similar to those of adults.

The similarity of adolescent and adult responses to trolley car situations has at least two major implications for research and theory on moral development. First, and most obviously, the finding suggests that multiple moral considerations about human life, and the ability to coordinate such considerations, emerge prior to adolescence. The development of these moral considerations poses intriguing puzzles for developmental research since most children do not have direct experiences with taking and saving lives. The development of these considerations may, therefore, be related to, for instance conversations with peers and family members, news reports, or works of fiction (Rosengren et al., 2014). A promising question for future research is how children use these and possibly other experiences regarding life and death to construct multiple moral concerns with human life.
Second, adolescents’ moral reasoning about life based on multiple considerations contrasts with the risk-taking or impulsivity often viewed as characteristic of adolescence (see Crone & Dahl, 2012). The present findings do not contradict findings of higher rates of risky behavior in adolescence. Yet, the present findings challenge any notion that adolescents engage in behaviors that put themselves or others at risk simply because they lack moral concerns with human life. As we discuss below, an important question for future research is how moral reasoning about human life and welfare is incorporated in the decision making of adolescents and adults.

It is possible that research using other situations could discover differences between how adolescents and adults reason about and evaluate acts of saving and sacrificing lives. The present findings imply that any such differences would not be due to a general lack of distinct moral concerns with human life among adolescents. Since adolescents appear to have these basic concerns, we would not expect developmental changes in responses to sacrificing and saving lives from adolescence to adulthood in response to straightforward situations involving a single concern, such as unprovoked killing. Rather, we would expect that developmental differences between adolescents and adults in moral reasoning to be evident mainly in certain multifaceted situations that prompt individuals to coordinate among competing concerns (e.g. Colby & Kohlberg, 1987a, 1987b). Indeed, as we noted in the introduction, it is primarily in responses to some multifaceted situations, not straightforward situations, that past research has found age-related differences in evaluations and reasoning in adolescence and beyond (Helwig, 1995; Nucci & Turiel, 2009; Turiel & Nucci, 2017).

The present work showed that, by adolescence, individuals do not approach the value of life in a uniform way. Rather, they have multiple moral concerns with the value of life that
sometimes conflict with each other, as in the standard trolley car situations. Past research has investigated how children coordinate moral concerns with others’ welfare with other moral and non-moral concerns, for instance equity or rights (Helwig, 1995; Nucci & Turiel, 2009; Rizzo & Killen, 2016). However, there has been less research on children’s developing abilities to deal with situations in which moral concerns with welfare or the value of life are in conflict with each other. A child’s construal and handling of conflicts between the lives and welfare of others will likely depend, in part, on the distinct concerns with welfare or life, for instance regarding victim involvement or rights to life, the child has acquired.

**Implications for Dual-Process Theories about Morality and its Development**

The use of trolley car situations has played a large role in recent research on adult moral psychology, which was one reason for using trolley car situations in the present research. In particular, past research on responses to trolley car situations have been central to dual-process accounts of moral psychology (Greene, 2013, 2014; Greene et al., 2001). According to these accounts, most moral evaluations are based on unconscious, affective reactions. In contrast, a smaller number of evaluations are said to be based on what researchers call “conscious reasoning” (see below for a critique of these terms).

The dual-process theories relate to theories of moral development in several ways. First, theories of moral development should be compatible with findings about adult morality, since developmental theories aim to account for how adult morality comes about. As discussed in the Introduction, research on moral developmental has documented that individuals reason about moral issues from an early age, which runs counter to the proposal that adults rarely reason about moral issues. Second, dual process theories often make developmental assumptions, for instance that certain affective reactions are innate (Greene, 2008; Rottman & Young, 2015; note that
Greene, 2017, states that he no longer believes that these affective reactions are innate. Conversely, some developmental researchers have taken dual-process claims about moral evaluations based on unconscious, affective reactions to support claims about innate moral evaluations in infants (Sloane, Baillargeon, & Premack, 2012; Van de Vondervoort & Hamlin, 2016). In short, claims about adults have implications for claims about moral development.

The findings reported in this Monograph present several challenges to dual-process accounts of morality. In subsequent sections we focus on two claims central to dual-process accounts of trolley car situations. The first claim challenged by the present findings is that people are typically unable to explain the reasons for their evaluations about the standard footbridge situations. The second claim we will discuss is that people are unconflicted about their evaluations in switch and even footbridge situations.

Participants provided reasons for their evaluations about both footbridge and switch situations. Psychologists studying responses to trolley car situations have long been concerned with whether people can provide reasons as to why it is wrong to intervene in the footbridge situation, but permissible to intervene in the switch situation (Cushman et al., 2006; Greene, 2014; Greene et al., 2001; Hauser et al., 2007). Based on previous research, dual-process theorists have argued that people’s evaluations of the footbridge situation typically do not involve reasoning, and that people are largely unaware as to why they evaluate the switch and footbridge situations differently (e.g. Cushman et al., 2006; Greene, 2014; Haidt, 2007; Hauser et al., 2007).

The claim that people typically cannot provide justifications for their evaluations of the standard footbridge situation has been particularly important for dual-process accounts (Greene, 2014; Haidt, 2007; Rottman & Young, 2015). For instance, Greene (2014, p. 705) writes:
“People are invariably conscious of the consequentialist rationale for making consequentialist judgments, but lack conscious access to the causes of their deontological patterns of judgment.” (By consequentialist, Greene means evaluations consistent with the utilitarian principle. By deontological, he means evaluations that are “naturally justified in deontological terms [in terms of rights, duties, etc.] and that are more difficult to justify in consequentialist terms, such as judgments against killing one person to save five others” [2014, p. 699].) As noted in Chapter 2, the claim about people’s inability to justify their evaluations about some trolley car situations is largely based on the studies by Cushman and colleagues (2006) and Hauser and colleagues (2007).

According to the dual-process view, the differing evaluations of the footbridge and switch situations are not based on reasoning about different morally relevant features of the situations. Instead, the dual-process view proposes that people say it is wrong to intervene in the footbridge situation due to an automatic emotional reaction, the basis for which is inaccessible to people. In contrast, the switch situation is said to elicit no such automatic emotional reaction, allowing participants to rely on reasoning about the number of lives lost and saved. The critical difference between the footbridge and switch situations, which is said to explain the different reactions to the footbridge and switch situations, is the “personalness/directness” and the intentionality of the act of pushing someone off a bridge (Greene, 2014; Greene et al., 2009). That is, the personal/direct and intentional nature of the act of pushing a person off a footbridge, as opposed to flipping a switch, is said to elicit an automatic emotional reaction, leading people to say pushing is wrong.

The findings of the present research show that people do provide reasons for why they think it is wrong to push the person off the footbridge but not wrong to intervene in the switch
variant. This finding challenges the dual-process claims that responses to footbridge situations are based on unconscious, automatic, and emotional reactions. In short, participants considered the footbridge and switch situations to differ in several ways that mattered for their evaluations, not just in the personalness and intentionality of the intervention. When explaining why it was wrong to intervene in the standard footbridge situation, participants referenced rights to life, the different involvement of the potential victims (bystander vs. workers on the track), and the rights and responsibilities of the potential intervener. All participants who judged intervention permissible in the switch but not footbridge situation provided at least one codeable explanation for why the situations differed from a moral point of view. Thus, there were no indications that participants could not reason about their evaluations of trolley car situations (For other challenges to claims about people’s inability to reason, see Jacobson, 2012; Royzman, Kim, & Leeman, 2015).

Moreover, the findings from the footbridge variant of the 1 scratched vs. 5 killed and the culprit vs. 5 workers situations indicate that people do not generally think the personalness and intentionality of pushing by itself makes it wrong. The vast majority of participants deemed that it was permissible to push a person off the footbridge when the person would merely be scratched and bruised or when it was a culprit, even though the physical act of pushing was the same as in the standard footbridge situation. Thus, personalness and intentionality of harm in the footbridge situation are not the only ways in which the standard footbridge situation differed from the switch situation; in fact, these dimensions did not appear to be the main considerations that guided people’s evaluations of pushing a person off the footbridge.

It appears that past research has overlooked types of justifications not falling into categories expected by the researchers. The utilitarian principle of maximizing welfare for the
greatest number is considered in some theories the main principle of moral reasoning (Greene, 2014), but was only one of several forms of reasoning invoked by participants. Past research on people’s responses to trolley car situations has not considered statements about the intrinsic rights and value of individuals, references to the person’s involvement in the situation, scope of responsibility, or considerations about guilt and social repercussions as indications of reasoning. A restrictive view of what counts as (moral) justifications may have prevented past research from documenting people’s reasoning about trolley car situations (Jacobson, 2012).

In addition to considering additional types of reasons, another distinctive feature of the present research was the use of follow-up questions to clarify people’s reasoning. Participants’ initial responses may rest on unstated premises and definitions that are only revealed through subsequent questioning. For instance, when asked why it was wrong to push the person off the footbridge in the standard version, one participant initially said: “Because he is innocent.” This response by itself may have given no indication as to why the participant judged intervention in the standard switch as permissible – are not all the potential victims in the standard trolley car situations innocent? However, the participants’ subsequent responses clarified his reasoning and made clear why this participant might view the switch situation differently: “Because he’s just an innocent bystander. […] His death would be a direct cost of my actions. Whereas if I didn’t push him, those five people who died would have died because of natural circumstances.” (This participant’s elaboration ultimately contained elements of the victim involvement, responsibility, and natural course justification categories, illustrating the complex considerations participants often articulated in response to interview questions.)

Previous studies of justifications about trolley car evaluations were limited by solely asking participants to explain differences between evaluations and not to explain their initial
evaluations (Cushman et al., 2006; Hauser et al., 2007). Justifying the difference between evaluations is different, and sometimes harder, than justifying the evaluations themselves. (As an analogy, we may compare the complex task of explaining whether Benito Mussolini or Augusto Pinochet was a worse dictator to the simpler task of explaining why either dictator was bad.) Nevertheless, Study 1 did find that participants justified their evaluations about differences between the standard variants. Past researchers may have failed to find such evidence by applying overly restrictive criteria for what counts as a justification of the difference between the switch and footbridge situations. For instance, Hauser and his colleagues (2007) considered as evidence of reasoning only justifications that “correctly identified any factual difference between the two scenarios and claimed the difference to be the basis of moral judgment” (p. 13). However, it can be problematic to determine whether a single statement provided in response to a single request for justification is identifying a factual difference, and even to determine what counts as factual differences. For instance, if a participant states that the person on the footbridge is an innocent bystander whereas the worker on the sidetrack is already in the situation, the participant would presumably say that he has identified a real and morally relevant difference between the footbridge and switch situation. But it is unclear from Hauser et al.’s (2007) coding description whether they would have considered such a statement the identification of a “factual difference.”

Another way justifications have been discounted in discussions of findings from trolley car studies is on the grounds that the justifications are said to be factually wrong (Cushman et al., 2006; Hauser et al., 2007). Examples of supposed factual inadequacies are justifications that included “added assumptions” not stated in the situation descriptions. For instance, Hauser and his colleagues (2007) would likely have excluded most victim involvement justifications because
such justifications included added assumptions about whether the workers had in fact accepted the risk of being on the track. But, as noted by Jacobson (2012) and Elster (2011), it is all but impossible to evaluate highly unusual (or even implausible) hypothetical situations without filling in background assumptions. How did the persons in the situation get there? What did the workers’ contracts look like? Have similar incidents happened before? How do you know that the man on the bridge will stop the train if pushed? Under these circumstances, the assumption that the workers have accepted the risks associated with railroad work is no less reasonable than the assumption that the workers have not accepted such risks.

These differences in interview and coding procedures likely explain why the present research found that people provided justifications for evaluations about a variety of trolley car situations, whereas past studies did not (Cushman et al., 2006; Hauser et al., 2007). Our findings are consistent with research on children’s, adolescents’ and adults’ moral evaluations and justifications in straightforward and multifaceted situations, which have shown that individuals make systematic evaluations based on welfare, justice, and rights (for reviews, see Killen & Smetana, 2015; Turiel, 2015). Adapting methods from past developmental research, the present studies involved standard questions and probes and found that these methods were applicable also to responses to trolley car situations.

Participants also reasoned about the number of lives saved. In this sense, our findings can be said to be partly consistent with claims of dual-process theories (Greene, 2014). However, we interpret participants’ utilitarian considerations differently than dual-process theories have, based on the present findings as well as the philosophical literature. That is, our findings did not indicate that people thought these utilitarian considerations should always be given priority over other considerations, such as personal responsibility for another person’s death. Moreover, in the
philosophical literature, the utilitarian principle is not generally accepted as the most valid, or only valid, principle for evaluating actions (Korsgaard, 1996; Wiggins, 2006; Williams, 1973). Some of the principles proposed by philosophers are reminiscent of the reasons provided by participants in our research. These include the right not to be used as a means to others’ ends (Kant, 1785; Thomson, 1976), or considerations regarding the integrity of the actor (Williams, 1973). Similarly, Foot, who constructed the initial trolley car situation (Foot, 1967), argued for the “existence of a morality which refuses to sanction the automatic sacrifice of the one for the good of the many … [and] secures to each individual a kind of moral space which others are not allowed to invade” (Foot, 1985, p. 36). As already noted, Williams (1973) argued against the utilitarian notion that all that matters, morally speaking, is the consequences of actions, not how those consequences are brought about. The assertions by Foot and Williams are clearly in contradiction of Greene’s (2013) proclamations about the moral irrelevance of directly causing someone’s death and align with reasons provided by participants in the present studies.

To further illustrate the ways moral concepts are based on more than utilitarian calculations regarding the greatest number of lives saved, we refer to a situation not yet mentioned (introduced by Foot, 1967, and used by Greene et al., 2001). It is a “transplant case” and not a trolley car situation. In the transplant case, the question posed is whether a surgeon can kill a healthy person in order to use the person’s organs to save the lives of five people who would otherwise die. Most philosophical treatments and the vast majority of research participants have deemed that it is not permissible for the surgeon to take the life of one person to save five lives. It seems that most people reject the validity of an utilitarian calculation in this case in favor of moral prescriptions bearing on the duties, roles, and responsibilities of a surgeon, and restricting the power granted by society and governments to individuals to make life and death
decisions – and, of course, the right to life of the healthy individual. Few would propose that people should not be sensitive to these features of the situation. This example shows that most people (including philosophers) would agree that there can be valid reasons for evaluations that go against simple utilitarian calculations based on the number of lives saved.

Further evidencing the relevance of non-utilitarian as well as utilitarian reasoning, participants’ justifications were consistent with their pattern of evaluations across situations. This indicates that the justifications participants provided highlighted situational features to which their evaluations were sensitive. Consistent with participants’ statements that victim involvement mattered, their evaluations were sensitive to victim involvement (Study 2). Consistent with statements that the responsibility of the protagonist mattered, participants said it was wrong to intervene even in the switch situation when the number of people sacrificed and saved were the same (*I killed vs. I killed*, Study 1). The justifications were part and parcel of their evaluations and not ways of rationalizing their decisions in unconnected ways (which some have referred to as post-hoc rationalizations). Similar consistencies between justifications, situational features, and evaluations were found for each of the justification categories used by participants, and have been found in a number of past studies with children and adults (Killen & Smetana, 2015; Turiel, 2015).

**Participants were conflicted about their evaluations, and did not view the standard switch or footbridge situations as straightforward.** According to dual-process accounts of trolley car situations, most participants should find the switch situation, and even the footbridge situation, relatively straightforward. In particular, people who say it is wrong to push the person off the footbridge are basing their evaluation on an automatic response, and Greene and his colleagues (2004) claimed that most participants found the footbridge situation “easy” (p. 395).
Similarly, people who say it is permissible to intervene in the switch situation are presumed to base this evaluation on utilitarian comparisons of the number of lives saved (Greene, 2014), which leaves little room for doubt (five is greater than one).

However, for both the switch and footbridge situations in Study 1, about 70 percent of participants expressed conflicts about their evaluations. These conflicts were indicated by dislike of the consequences of their overall evaluation, consideration of arguments for the opposing evaluation, and accepting of alternative evaluations. Rates of conflicts were far higher in the standard situations (1 killed vs. 5 killed) than in the expectably less conflictful situation in which the person to be sacrificed was only scratched and bruised (1 scratched vs. 5 killed).

The widespread presence of conflict is difficult to explain from the point of view of dual-process accounts. While the latter accounts expect participants to be conflicted when deeming that it is permissible to intervene in the footbridge situation, they yield no obvious explanation for why people would be conflicted about the switch situation. In the dual-process view, the switch situation should not elicit aversion to pushing or any other automatic, affective reactions that could compete with the utilitarian considerations about the number of lives saved.

As noted, the evidence of conflict in the standard switch and footbridge situations is consistent with our proposition that the standard trolley car situations elicit reasoning about distinct and competing moral concerns with the value of life. In our account, conflicts ensue in people’s reasoning about these situations because they are uncertain about or discontent with their prioritization of some concerns, such as the number of lives saved, over others, for instance considerations about personal responsibility and the right to life (Białek & Neys, 2016; Williams, 1973).
Reasoning, Emotions, and Actions beyond Trolley Cars: Moral Orientations about Sacrificing and Saving Lives

We view reasoning as central to most aspects of morality. Importantly, we adopt a different definition of reasoning and different methods for studying reasoning than dual-process views and related theories. In our view, reasoning often involves coordination of competing considerations (for instance about trolley car situations) and is closely intertwined with emotional reactions and actions. In this section, we briefly discuss each of these issues (for further discussion, see Dahl & Killen, 2017; Turiel, 2015; Turiel & Dahl, in press).

Evidence for reasoning about moral issues. In the Introduction, we defined moral reasoning as the formation of moral evaluations based on moral considerations that people can endorse and articulate. Assessing whether people reasoned about a given issue typically involves two related inquiries. First, it involves asking individuals to justify their evaluations. Second, it involves investigating whether the concerns referenced in individuals’ justifications are reflected in their evaluations and justifications across situations. In particular, are people’s evaluations sensitive to the situational features referenced in their justifications? The present research included several examples of this procedure, and provided evidence that people’s evaluations were indeed sensitive to the situational features referenced in their justifications.

As we noted in the Introduction, we are not thereby assuming that individuals are deliberately going through steps of reasoning. Philosophical and psychological accounts of reasoning in other domains view reasoning as proceeding at different speeds, and the presence of reasoning does not hinge on being consciously aware of transitions in thought (Adler & Rips, 2008; Harman, 1986). In our view, moral reasoning may sometimes happen slowly and deliberatively and other times rapidly and effortlessly, similar for instance to reasoning about
mathematics or physics (e.g. Larkin, McDermott, Simon, & Simon, 1980; Pizarro & Bloom, 2003). We do not assume that people consciously go through steps prior to making an evaluation when they engage in moral reasoning.

We are not proposing that reasoning fully determines every aspect of morality. For instance, inducing disgust in participants can have small but significant effects on ratings of others’ actions (for a review, see Landy & Goodwin, 2015), such as moving participants’ ratings of others’ actions from 55 to 65 on a quantitative rating scale from 0 (“not at all wrong”) to 100 (“extremely wrong”, see e.g. Wheatley & Haidt, 2005). We suspect that participants will not always be aware of, or provide justifications for, such small shifts in quantitative ratings of actions. However, participants may also not think that such shifts are morally significant, and may find it acceptable that another person rated an action slightly differently than them (Banks, 2016). Our claim is that reasoning matters greatly for moral evaluation, but not that reasoning is the only thing that matters.

According to this definition of moral reasoning, the findings of the present research support the proposition that people from adolescence to adulthood make evaluations about a variety of life and death situations based on reasoning. This conclusion is consistent with findings from a large number of studies with children, adolescents, and adults (Turiel, 2015). Our conceptualization of reasoning differs from that of dual-process theories. For instance, Greene and Paxton define moral reasoning as “[c]onscious mental activity through which one evaluates a moral judgment for its (in)consistency with other moral commitments, where these commitments are to one or more moral principles and (in some cases) particular moral judgments” (Paxton & Greene, 2010, p. 516). A fundamental claim of dual-process approaches is
that moral evaluations are based either on such conscious moral reasoning or – more frequently – on automatic, unconscious reactions (“intuitions”) (Greene, 2014; Paxton & Greene, 2010).

In our view, there are several problems with the categorization of moral evaluations as based on either (conscious) reasoning or automatic, unconscious intuitions. The first problem is that the definition of reasoning as conscious conflicts with common definitions of reasoning. These common definitions do not restrict reasoning to conscious activity (Adler & Rips, 2008; Harman, 1986; Harman, Mason, & Sinnott-Armstrong, 2010). Moreover, standard accounts of reasoning do not limit reasoning to assessing consistency between a particular statement and a general commitment, but also include other forms of reasoning, for instance deducing a conclusion from premises. Thus, the dual process models of moral evaluations use “reasoning” in a highly unusual way. The second problem with the dual-process definition of reasoning is that past research on moral evaluations has typically not sought to assess whether participants are engaging in moral reasoning as defined by Paxton and Greene (2010; see Kihlstrom, 2008, 2013). The only method that comes close is the method of asking participants to justify their answers, which we discussed above. However, not even the interview method is directly assessing whether participants had conscious awareness of going through steps prior to making an evaluation, since participants are only asked to justify their evaluations after making the evaluation. Without assessing whether people engage in reasoning in the sense proposed by Greene and Paxton (2010), it is also impossible to determine whether all moral evaluations are based on either conscious reasoning or automatic, unconscious reactions, or whether there other types of evaluations, for instance evaluations based on both controlled and automatic processes.

We adopted our alternative definition both based on philosophical and psychological treatments of reasoning. This definition has the added benefit of avoiding the notorious difficulty
of seeking to determine individuals’ conscious awareness about the formation of an evaluation. If the necessary methods become available, assessing variations in levels of conscious awareness will be of interest. However, in our view, the major question regarding moral reasoning is whether people can access and endorse the considerations on which their evaluations are based: Do people know their reasons for evaluating some actions as right and others as wrong, and do they view those reasons as valid? With respect to participants’ responses to trolley car situations the answer based on the present findings was “yes.”

**Coordination in reasoning about sacrificing and saving lives.** In our view, moral reasoning implies that people are committed to a given moral concern across situations. However, this does not mean that people will always give priority to the same concerns in all situations. In multifaceted situations, moral reasoning will sometimes involve conflict, coordination, drawing priorities, and uncertainty (Turiel, 2008a). Conflict and coordination is not unique to reasoning about moral issues. In physics, scientists famously reasoned and researched for decades in attempts to resolve the apparent contradiction between viewing light as particles and viewing light as waves (Einstein & Infeld, 1938). Adherence to conflicting statements does not, by itself, indicate irrationality or an absence of reasoning, in morality or in physics.

We view reasoning about conflicting concerns as central to people’s responses to trolley car situations, as well as other multifaceted moral situations. We have maintained that situations like the trolley car problems present people with conflicting components requiring processes of coordination. We do not view the general prohibition against killing as an automatic reaction to the act of pushing, but as understandings that people adhere to in most situations and are only willing to violate under extreme circumstances (e.g. when necessary to save other lives, or in self-defense). The findings of our studies point to the salience of processes of thought...
involved in making decisions about the trolley car situations. In these types of situations that entail strong conflicts, such as the conflict posed by needing to violate the value of life to preserve that very value, there is need to recognize that decisions require weighing and balancing different considerations.

The proposition that decisions can involve attempts to coordinate conflicting considerations is illustrated by responses to several of the footbridge variants in Study 1. On our account, most people say it is wrong to push another person off the footbridge in the standard variant because they prioritize obligations toward person of the footbridge over obligations toward the five persons on the track, not because of an emotional reaction to the act of pushing another. Several non-standard footbridge variants showed that participants weighed both the consequences to the person to be pushed and the nature of the person’s on the track when evaluating footbridge situations. Most participants accepted pushing the person off the bridge if the person would only be scratched and bruised (but not killed) or if the five persons to be saved were family members. If the mental image of pushing a person was sufficient to make people condemn intervention in the footbridge situation, participants should have judged intervention as impermissible certainly in the 1 vs. 5 family situation and even in the 1 scratched vs. 5 killed situation. On the contrary, the pattern of evaluations indicated that participants evaluated acts of pushing to largely based on the consequences of the pushing.

Participants’ efforts to coordinate competing considerations were also reflected in their expressions of conflict, which were equally common in the standard switch and footbridge situations. The data on conflict show that participants did not approach the switch situation in purely utilitarian terms. As noted above, most participants did not view the standard switch situation as straightforward. One participant expressed the following doubts about saying it was
alright to intervene in the standard switch situation: “I think life is precious and it would be a very sad thing. [...] I don’t think it would be okay for the [one] person to die, but I would feel worse if more than one person died and I could have done something about it.” This quote illustrates a general pattern in our data: in most cases, participants viewed the trolley car situations as requiring them to balance competing principles about right and wrong in order to form an evaluation.

Furthermore, when asked whether it was permissible to sacrifice one person to save a single worker (1 killed vs. 1 killed), over 90% of participants said it was not. If participants had viewed the two possible outcomes in the switch situation purely in terms of number of lives lost they would have viewed the outcomes as morally equivalent (one person dies either way). However, participants saw major differences between intervention and non-intervention, referring to the intrinsic value and rights of individuals and the natural course of events in explaining why it was not permissible to flip the switch to sacrifice one person to save another person.

In summary, the present research on reasoning about sacrificing and saving lives demonstrate that people’s moral orientations toward the value of life is heterogeneous. Adolescents and adults indicated that human lives are valuable and that individuals have obligations toward the lives and welfare of others. However, orientations toward the value of life contained multiple moral concerns that sometimes entered into conflict with each other. Thus, while most people find it straightforward to say that we ought to protect human lives, they may not find it straightforward to determine the right way of protecting human lives in multifaceted situations.
Moral reasoning, emotion, and action. Thus far in our account of people’s orientations to saving and sacrificing lives we have emphasized the role of reasoning. This is in part because of the central role we give to reasoning in morality, but also because the present study was focused on people’s evaluations and justifications. However, we do not view reasoning as independent of emotions. On the contrary, we view reasoning and emotional processes as operating in concert, not in opposition to each other (Turiel & Dahl, in press; Turiel & Killen, 2010). Our basic proposition regarding relations between emotions and morality is that people can be emotional about any moral issue. Emotional reactions accompany moral evaluations and reasoning when individuals view events as relevant to the fate of their moral concerns (Frijda, 1986). The close intertwining of thoughts and emotions are consistent with the view, held by many emotion theorists, that most or all emotional reactions are constituted by cognitive appraisals (Lazarus, 1991; Moors & Scherer, 2013; Nussbaum, 2001). According to this perspective, emotions are elicited when individuals appraise events as relevant to their concerns, such as moral concerns. Our proposal that reasoning and emotions are intertwined is also largely consistent with the views of Piaget (1981) and Kohlberg (1971), who both argued that affect and cognition were features of the same mental phenomena. (See Russell, 2014, for the related view that all experiences have an affective component.)

In our view, emotional reactions do not preclude reasoning, nor does reasoning preclude emotional reactions. Instead, we propose that there is a great deal of intersection between reasoning and emotional reactions, such that when people deem that an action is morally wrong they are also prone to have a negative emotional reaction to the event. For instance, individuals often negatively evaluate aggressive or harmful actions toward others, and have negative emotional reactions toward aggressive or harmful actions (e.g. Bonanno & Keltner, 2004;
Moral reasoning

Cushman et al., 2012; Dahl et al., 2014; Hoffman, 2000). If people are conflicted about how to evaluate an action, we would also expect them to have mixed emotional reactions to the action.

Our view of the connections between moral reasoning and emotions differs from the dual-process views of morality. Dual-process theories typically characterize reactions presumed to be automatic and unconscious as “emotional.” For instance, Greene writes that “moral judgment is influenced by both automatic emotional responses (automatic settings) and controlled, conscious reasoning (manual mode)” (2014, p. 698). As discussed earlier, such automatic emotional reactions are said to bring about non-utilitarian evaluation patterns. Importantly, most dual-process views also acknowledge that cognitive processes are involved even in the formation of presumed automatic evaluations (Greene, 2008, 2014; Greene et al., 2004; Mikhail, 2007).

The claim that emotions are exclusively linked to automatic, unconscious reactions conflicts with both empirical evidence and theories about emotions. As we stated in the Introduction, there is indeed no evidence that non-automatic processes generally occur in the absence of emotional arousal. At most, what has been shown is a relative difference in emotional activation in response to footbridge situations, presumed to elicit automatic evaluations, compared to switch situations, presumed to elicit reasoning (Greene et al., 2004, 2001). Furthermore, at least one study has found evidence that people do indeed become emotional about the standard switch situation (Skulmowski et al., 2014). Moreover, emotion theories hold that people could become emotional about any moral issue, not just non-utilitarian considerations, provided people care enough about the issue (Barrett & Campos, 1987; Frijda, 1986; Lazarus, 1991; Moors & Scherer, 2013). Perhaps reflecting these challenges, dual-process theorists sometimes acknowledge that people can be emotionally aroused about any issue, not
just non-utilitarian considerations. For instance, Greene has argued that “all moral judgment (including consequentialist judgment) must have some emotional component” (Greene, 2008, p. 41). Some theorists have indicated that they have called automatic, unconscious reactions “emotional” for reasons of “euphony” rather than clarity (Haidt & Bjorklund, 2008b, p. 250). Yet, despite some acknowledgement that emotional arousal can accompany any moral evaluation, current presentations of dual-process models continue to describe automatic unconscious reactions, in contrast to reasoning processes, as emotional (Greene, 2014; Haidt, 2008; Hayakawa et al., 2017; Rottman & Young, 2015).

Rather than contrasting emotion with reason, we propose that people can reason and have emotions about any moral issue. Intense conflicts regarding moral and non-moral issues are not constituted by a battle between emotions and reasons, but between emotions and reasoning about one consideration and emotions and reasoning about other, competing considerations. Although emotions have sometimes been associated with unreasonableness or irrationality in the media, among laypersons, and, historically, among scholars, contemporary emotion theories have largely abandoned this construal (Lazarus, 2001).

In proposing a close connection between reasoning and emotion, we are also proposing a close connection between reasoning and action. Insofar as people are concerned with moral issues, like the value of life, we expect them to reason about these issues, have emotional reactions about these issues, and be inclined to act to protect and promote their concerns (Frijda, 1986; Lazarus, 1991; Turiel, 2003). For instance, most people think it is generally wrong to harm others, have negative emotional reactions to harmful actions, and refrain from harming others in most situations.
In investigating relations between reasoning, evaluations, and action, it is necessary to consider the situations people reason about and in which they act. Studies reporting low relations between evaluations, reasoning, and action often assess evaluations and reasoning about one type of situation and action in a different type of situation. Relatedly, these studies rarely consider how individuals construe those situations, which may differ from how researchers construe the situations (Turiel, 2003). A further complication is that people sometimes act in situations posing conflicts, just as they sometimes reason about situations of conflict (Turiel & Dahl, in press). If participants are conflicted about what constitutes the right action, there may be a low correlation between evaluations of a particular kind and actions because people are considering different courses of action. As seen in many studies of decisions regarding hypothetical situations, there is conflict about different considerations and a process of coordination (Turiel, 2015). Similar processes of conflict and coordination occur in people’s behaviors, as seen in several social psychological experiments (Turiel, 2002). A good example can be seen in the courses of actions within Milgram’s (1974) experiments on so-called obedience to authority. As detailed elsewhere (Turiel, 2002), participants were not simply obeying authority. Rather, they were dealing with a choice between adhering to the scientific goals of the experiment and inflicting pain on another person. As is well known, participants were instructed to administer electric shocks in the guise of a study on learning and memory. The behaviors of most throughout the experimental session revealed a great deal of conflict experienced connected to their desire not to inflict pain. The conflicts generally occurred regardless of the ultimate decision, which involved drawing priorities between the two goals.

Adolescents and adults are unlikely to find themselves on a footbridge or a switch faced with the same situation as the protagonist in the trolley car situations. If they did, the physical
and psychological difficulties of flipping a switch and pushing a person off a footbridge, along with the uncertainty of whether the intervention would actually save the five persons, further complicate the relation between reasoning, evaluations, and actions. However, many individuals do find themselves in situations where their actions can make a difference for the survival of others. These situations range from extreme examples of saving individuals during Holocaust (Oliner & Oliner, 1988), to voting or expressing one’s opinion on capital punishment or abortion, to making potentially life-saving donations to humanitarian organizations.

To understand people’s decisions and actions in both extreme and everyday situations, the approach taken in the present research will likely be productive. In these situations, we anticipate that individuals will strive to coordinate and draw priorities among distinct and sometimes competing concerns, including concerns with the life and welfare of themselves and others. Sometimes, individuals may coordinate these concerns in ways that leave them with a clear conviction about a right course of action, which we anticipate will make them likely to act accordingly. Other times, individuals may remain conflicted about what they view as the right course of action. Based on the present findings, we expect such conflicts and hesitations to be reflected in how individuals approach, talk about, and act when faced with choices of taking or saving lives.

Conclusion

The present studies investigated people’s evaluations and justifications about several trolley car situations. The findings revealed that, by adolescence, people reason about such situations: Participants’ evaluations and justifications were sensitive to variations in situational features, and their evaluations involved attempts to coordinate conflicting principles, rather than a straightforward counting of lives saved. As a whole, the study documents the centrality of
reasoning when people evaluate possible actions in trolley car situations. The findings of this study can be interpreted in the context of the patterns of reasoning documented in a large number of studies on moral development. The systematic forms of thinking about morality as non-contingent on rules or authority and as not defined by common practices are in keeping with underlying conceptions of welfare, justice, fairness, and rights. The multifaceted situations used in our research led participants to reason about the value of life based on numbers and intrinsic rights as well as the responsibilities of the potential victims and the person intervening. Although the trolley car situations are unusual, participants thus perceived the situations to include significant moral components. The research presented here provides an alternative to the proposition that many moral evaluations are based on affective and automatic responses with little connection to systematic reasoning.

This research highlights the value of rooting developmental and other psychological research in philosophical analyses (Kohlberg, 1968, 1971; Piaget, 1965; Turiel, 2010). The philosophical literature furnished definitions and arguments that guided the present work. The trolley car situations themselves originated in the philosophical literature. Despite being unusual and even unrealistic, the situations successfully elicited distinct forms of reasoning about human life across the different groups of participants. Our hypotheses that participants would reason about both utilitarian and non-utilitarian considerations were informed by philosophical discussions of trolley car situations and critiques of utilitarianism (Foot, 1967; Thomson, 1985, 2008; Wiggins, 2006; Williams, 1973).

Nevertheless, there are crucial differences between individuals engaging in moral reasoning and philosophers proposing theories of morality. Alluding to Kohlberg’s (1968) notion of “The child as a moral philosopher,” we may say that children, adolescents, adults, and moral
psychologists in an important sense are not like moral philosophers. Differences between these groups are seen when considering how they deal with internal conflicts.

For moral philosophers, conflicting statements are typically a basis for criticism, indicating inadequacy of the philosophical theses put forth. Traditionally, philosophers aim to propose philosophical systems that are internally consistent, such that no one assertion conflicts with other assertions (for alternative approaches, see e.g. Wittgenstein, 1953). For instance, if a philosopher puts forth a utilitarian philosophy in which it is required to maximize the number of lives saved, yet argues that it is not permissible to push one person off the footbridge to save the five workers, the philosopher would be vulnerable to criticisms of being inconsistent.

In contrast, for an individual engaged in moral reasoning, conflicts between competing considerations are part of moral life. A person may hold that (A) other things being equal it is better to save five lives than to save one life and (B) other things being equal it is better not to bring about someone’s death. The problem, of course, is that “other things” are not always equal, sometimes leaving individuals without easy ways of resolving conflicts between their evaluative considerations. The signs of conflict in participants’ responses to the standard footbridge and switch situations indicate such unresolved tensions in participants’ moral orientations. However, unlike contradictions within moral philosophies, conflicts in the moral orientations of individuals do not indicate a shortcoming, nor do they indicate that individuals are irrational or operate on a lower stage of moral reasoning (Turiel, 2008b). Conflicts may instead be the consequences of relations among the distinct moral and non-moral concerns that individuals have developed. People’s moral concerns with life and welfare yield simple answers in most straightforward situations, as when evaluating acts of unprovoked violence, but clash in some multifaceted situations, as when evaluating the act of sacrificing one life to save others.
This Monograph dealt with how adolescents and adults reason about conflicting moral concerns in the context of sacrificing and saving lives. The findings challenge the notion that a person’s moral reasoning is characterized by a single, coherent orientation, such as utilitarianism. While their responses in structured interviews thus differed from the theories of moral philosophers, adolescents and adults reasoned systematically about complex situations designed to shed light on tensions among moral concerns with human life.
References


criteria and justifications in children’s social reasoning. *British Journal of Developmental


NY: Knopf.


https://doi.org/10.1016/j.cognition.2016.07.002


5–15.

(pp. 23–38). London: Routledge.

Angeles: SAGE Publications.


*Developmental Psychology, 44*, 875–881. https://doi.org/10.1037/0012-1649.44.3.875


*Judgment & Decision Making, 10*, 296–313.


https://doi.org/10.2307/796133


Acknowledgements

This research was supported in part by a grant from the National Institute of Child Health and Human Development (R03HD087590) to A. Dahl. We thank the research assistants who transcribed the in-person interviews, Lindsay M. Banks and Celina Kasprick for coding the interviews, and Will Torrey for comments on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Audun Dahl, Department of Psychology, University of California, Santa Cruz, CA 95064. E-mail: dahl@ucsc.edu.
Contributors

Audun Dahl is an Assistant Professor in the Department of Psychology at the University of California, Santa Cruz. His research investigates the development and application of fundamental moral concerns with human welfare from infancy to adulthood.

Matthew Gingo is an Assistant Professor in the Department of Psychology at Wheaton College. His research deals with the development of reasoning about social subversion and moral resistance in children and adolescents.

Kevin Uttich is a Senior Researcher at Conifer Research in Chicago, IL. He received his Ph.D. in Psychology from the University of California, Berkeley, where he studied the connections between social and moral cognition in children and adults.

Elliot Turiel is the Jerome A. Hutto Professor of Education at the University of California, Berkeley. His research is on social and moral development from childhood to adulthood. He also studies social opposition to cultural practices of inequalities.
### Tables

**Coding Categories**

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of life based on numbers</td>
<td>Consideration of the number of lives and/or sum of welfare involved.</td>
<td>“It’s better for one person to die than for five persons to die.”</td>
</tr>
<tr>
<td>Value of life based on rights</td>
<td>Reference to people’s intrinsic rights or values.</td>
<td>“You’re violating someone’s rights by killing them.”</td>
</tr>
<tr>
<td>Victim involvement</td>
<td>Obligations to the role and contractual expectations of the person.</td>
<td>“The guy on the bridge had nothing to do with the situation/”</td>
</tr>
<tr>
<td>Natural course of events</td>
<td>Statement about non-interference in the natural course of events (or “fate”).</td>
<td>“That’s what was going to happen if I wasn’t there to intervene.”</td>
</tr>
<tr>
<td>Responsibility for consequences of actions</td>
<td>Agent would be responsible for outcome.</td>
<td>“Then I’d be responsible for that person’s death.”</td>
</tr>
<tr>
<td>Consequences to self</td>
<td>References to consequences for agent, e.g. distress, self-criticism, or being negatively evaluated by others.</td>
<td>“I would feel horrible if I intervened.”</td>
</tr>
</tbody>
</table>
Table 2

*Study I: Percent of Participants Responding it was Permissible to Intervene*

<table>
<thead>
<tr>
<th>Situation</th>
<th>Group</th>
<th>Group</th>
<th>Group</th>
<th>vs. Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adolescent in-person</td>
<td>Student in-person</td>
<td>Student online</td>
<td>Adult online</td>
</tr>
<tr>
<td>Standard (1 vs. 5)</td>
<td>Footbridge</td>
<td>19</td>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>71</td>
<td>73</td>
<td>55</td>
</tr>
<tr>
<td>1 scratched vs. 5</td>
<td>Footbridge</td>
<td>92</td>
<td>92</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>100</td>
<td>96</td>
<td>89</td>
</tr>
<tr>
<td>1 vs. 5 family</td>
<td>Footbridge</td>
<td>54</td>
<td>38</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>83</td>
<td>83</td>
<td>78</td>
</tr>
<tr>
<td>Dog vs. 5</td>
<td>Footbridge</td>
<td>92</td>
<td>83</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>100</td>
<td>96</td>
<td>80</td>
</tr>
<tr>
<td>1 vs. 1</td>
<td>Footbridge</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>04</td>
<td>04</td>
<td>10</td>
</tr>
</tbody>
</table>

*Note.* Rightmost column shows significance of difference from corresponding standard variant (e.g. 1 killed vs. 1 killed footbridge vs. standard footbridge), as assessed with likelihood ratio tests comparing GLMMs. *p < .05, **p < .01, ***p < .01.*
Table 3

*Study 1: Justifications for Evaluations in Standard Situations*

<table>
<thead>
<tr>
<th>Justification</th>
<th>Percent of justifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of life based on numbers</td>
<td>27</td>
</tr>
<tr>
<td>Value of life based on rights</td>
<td>6</td>
</tr>
<tr>
<td>Victim involvement</td>
<td>14</td>
</tr>
<tr>
<td>Natural course of events</td>
<td>18</td>
</tr>
<tr>
<td>Responsibility for consequences of actions</td>
<td>24</td>
</tr>
<tr>
<td>Consequences to self</td>
<td>11</td>
</tr>
</tbody>
</table>

*Note.* Table shows proportion of coded justifications for evaluations about Standard situations falling into each category (N = 288 participants).
### Table 4

**Study 1: Justifications by Situation, Evaluation, and Group (Standard Situations)**

<table>
<thead>
<tr>
<th>Situation</th>
<th>Evaluation</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adolescent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in-person</td>
</tr>
<tr>
<td><strong>Value of life based on numbers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>100</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>100</td>
</tr>
<tr>
<td><strong>Value of life based on rights</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>11</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>29</td>
</tr>
<tr>
<td><strong>Victim involvement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>18</td>
</tr>
<tr>
<td><strong>Natural course of events</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>15</td>
</tr>
<tr>
<td><strong>Responsibility for consequences of actions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>11</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>9</td>
</tr>
<tr>
<td><strong>Consequences to self</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>11</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>21</td>
</tr>
</tbody>
</table>

*Note.* Cells show percent of participants making a given evaluation in a given situation who provided a given justification (Standard variants only, $N = 288$ participants). For instance, the bottom right cell shows that among Adult (online) participants who said it was alright to intervene in the switch situation, 14% used a *consequences to self* justification.
Table 5

Study 1: Justifications for “Typical” Evaluation Combination (Standard Switch Permissible, Standard Footbridge not Permissible)

<table>
<thead>
<tr>
<th>Situation</th>
<th>Group</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Value of life, numbers</td>
</tr>
<tr>
<td>Footbridge</td>
<td>Adolescent in-person</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Student in-person</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Student online</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Adult online</td>
<td>13</td>
</tr>
<tr>
<td>Switch</td>
<td>Adolescent in-person</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Student in-person</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Student online</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Adult online</td>
<td>13</td>
</tr>
</tbody>
</table>

Note. Cells show percent of participants who, in explaining why they judged the standard footbridge and trolley situations differently, used a given justification with respect to a given situation (N = 118 participants).
Table 6

Study 1: Expressions of Conflict (Standard Situations)

<table>
<thead>
<tr>
<th>Situation</th>
<th>Group</th>
<th>Stating that consequences are “not alright”</th>
<th>Considering arguments for contrary evaluation</th>
<th>Accepting opposite decision by someone else</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adolescent in-person</td>
<td>Student in-person</td>
<td>Student Online</td>
<td>Adult online</td>
</tr>
<tr>
<td>Footbridge</td>
<td>21</td>
<td>42</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>Switch</td>
<td>21</td>
<td>41</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contingency</td>
<td>Student In-Person</td>
<td>Student Online</td>
<td>Adult online</td>
</tr>
<tr>
<td>Footbridge</td>
<td>48</td>
<td>53</td>
<td>68</td>
<td>52</td>
</tr>
<tr>
<td>Switch</td>
<td>67</td>
<td>50</td>
<td>60</td>
<td>56</td>
</tr>
</tbody>
</table>

*Note.* The cells show percent of participants who expressed the different signs of conflict as a function of situation and group (N = 288 participants).
Table 7

*Study 1: Justifications for Contrary Evaluations (Standard Situations)*

<table>
<thead>
<tr>
<th>Situation</th>
<th>Participant’s own evaluation</th>
<th>Justification for Contrary Evaluation</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value of life, own evaluation</td>
<td>Value of life, numbers</td>
<td>Value of life, rights</td>
<td>Victim involvement</td>
<td>Natural course of events</td>
<td>Responsibility for consequences</td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>75</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>0</td>
<td>38</td>
<td>25</td>
<td>0</td>
<td>62</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>64</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>4</td>
<td>29</td>
<td>17</td>
<td>17</td>
<td>12</td>
</tr>
</tbody>
</table>

*Note.* Cells show percent of participants making a given evaluation for a given standard situation who discussed a justification for the opposite evaluation ($N = 288$ participants).
Table 8

*Study 1: Use of Justifications in Non-Standard Situations*

<table>
<thead>
<tr>
<th>Situation</th>
<th>Evaluation</th>
<th>Value of life, numbers</th>
<th>Value of life, rights</th>
<th>Victim involvement</th>
<th>Natural course of events</th>
<th>Responsibility for consequences</th>
<th>Consequences to self</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Footbridge</th>
<th>Ok</th>
<th>1 scratch vs. 5 killed</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>54</td>
<td>25</td>
<td>14</td>
<td>43</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>82</td>
<td>86</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>9</td>
<td>64</td>
<td>0</td>
<td>27</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>83</td>
<td>85</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Eval. effect</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Situation effect</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Group effect</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Footbridge</th>
<th>Ok</th>
<th>1 killed vs. 5 family killed</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>70</td>
<td>30</td>
<td>22</td>
<td>49</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>81</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>11</td>
<td>69</td>
<td>11</td>
<td>33</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>58</td>
<td>4</td>
<td>15</td>
<td>4</td>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Eval. effect</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Situation effect</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Group effect</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>
### Justifications

<table>
<thead>
<tr>
<th>Situation</th>
<th>Evaluation</th>
<th>Value of life, numbers</th>
<th>Value of life, rights</th>
<th>Victim involvement</th>
<th>Natural course of events</th>
<th>Responsibility for consequences</th>
<th>Consequences to self</th>
</tr>
</thead>
<tbody>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>8</td>
<td>58</td>
<td>29</td>
<td>11</td>
<td>50</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>95</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>5</td>
<td>76</td>
<td>14</td>
<td>24</td>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>95</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

**Eval. effect**
-***

**Situation effect**
-***

**Group effect**
-***

### 1 dog killed vs. 5 people killed

<table>
<thead>
<tr>
<th>Situation</th>
<th>Evaluation</th>
<th>Value of life, numbers</th>
<th>Value of life, rights</th>
<th>Victim involvement</th>
<th>Natural course of events</th>
<th>Responsibility for consequences</th>
<th>Consequences to self</th>
</tr>
</thead>
<tbody>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>23</td>
<td>58</td>
<td>37</td>
<td>52</td>
<td>40</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>22</td>
<td>60</td>
<td>17</td>
<td>67</td>
<td>26</td>
<td>9</td>
</tr>
</tbody>
</table>

**Eval. effect**
- -

**Situation effect**
-***

**Group effect**
-***

**Note.** Cells show percent of participants who provided each justification to justify the given evaluation in the given situation (N = 288 participants). The bottom three rows for each situation indicate statistical significance levels for likelihood ratio tests of evaluation, situation, and group effects using binomial GLMMs (“-“ indicates that predictor was not included in model because too few participants said intervention was permissible.). *p < .05, **p < .01, ***p < .001.
Table 9

*Study 2: Percent of Participants Saying it was Permissible to Intervene*

<table>
<thead>
<tr>
<th>Situation</th>
<th>Culprit vs. 5</th>
<th>Worker vs. 5</th>
<th>Non-worker vs. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Footbridge</td>
<td>73</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Switch</td>
<td>92</td>
<td>65</td>
<td>52</td>
</tr>
</tbody>
</table>

*Note.* Cells indicate, for each situation, the percent of participants who said it was permissible to intervene (*N* = 144 participants).
### Table 10

**Study 2: Use of Justifications**

<table>
<thead>
<tr>
<th>Situation Pair</th>
<th>Evaluation</th>
<th>Culprit vs. 5</th>
<th>Worker vs. 5</th>
<th>Non-worker vs. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Value of life based on numbers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>7</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>64</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>67</td>
<td>100</td>
<td>88</td>
</tr>
<tr>
<td><strong>Value of life based on rights</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>40</td>
<td>61</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>67</td>
<td>65</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>9</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Victim involvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>20</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>88</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>0</td>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>78</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td><strong>Natural course of events</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>20</td>
<td>34</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>3</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>0</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>2</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td><strong>Responsibility for consequences of actions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>47</td>
<td>68</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>6</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>0</td>
<td>53</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td><strong>Consequences to self</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>Not ok</td>
<td>27</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>Not ok</td>
<td>33</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Ok</td>
<td>7</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note.* Cells show proportions of participants making a given evaluation in a given situation who provided a given justification (N = 144 participants).