UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

THE GOOD CHILD, THE VIRTUOUS PROFESSOR AND ME: MORAL JUDGEMENT IN A SOCIAL CONTEXT

André Ricardo Amaral

MESTRADO INTEGRADO EM PSICOLOGIA

Área de Especialização em Cognição Social Aplicada

2020

UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

THE GOOD CHILD, THE VIRTUOUS PROFESSOR AND ME: MORAL JUDGEMENT IN A SOCIAL CONTEXT

André Ricardo Amaral

Dissertação orientada pelo Professor Mário Ferreira

MESTRADO INTEGRADO EM PSICOLOGIA

Área de Especialização em Cognição Social Aplicada

2020

Agradecimentos

Antes de mais, gostaria de agradecer ao meu orientador pela paciência dispensada, pela confiança e autonomia que me auferiu e por tudo o que aprendi. Sinto que saio desta tese mais rico, por ter tido o privilégio de trabalhar com o Professor Mário Ferreira.

Queria agradecer também aos professores deste Núcleo pela sua ajuda e dedicação incríveis e pelo seu gosto em transmitir conhecimento e fomentar a investigação e o gusto pela Ciência. Em especial ao Professor Tomás Palma, ao Professor José Palma, ao Professor André Mata e ao Professor Doutor Leonel Garcia-Marques.

Gostaria de agradecer também ao grupo de estagiários do Mestrado em Cognição Social aplicada, um verdadeiro *focus group* com capacidade e ideias intermináveis. Um agradecimento em especial à minha nobre parceira de trabalho, Mafalda Catarino, às horas que passámos juntos a trabalhar, a discutir ideias ou simplesmente a inserir dados. Foi de uma ajuda e de uma motivação inestimável e saio deste mestrado agradecido por esta amizade que sei que durará pela vida fora.

Depois, gostaria de agradecer à minha família e amigos que, estando longe, nunca deixaram de apoiar as minhas escolhas académicas e que sempre forneceram uma perspetiva que não tinha considerado sobre tanta coisa, desde a mais ínfima preocupação à mais importante metodologia de investigação.

Um especial agradecimento também à minha segunda família, Ana, Bruno, Daniela e Edna; e ao G.A.B. e seus colaboradores, Inês, Ana, Patrícia, Carolina, Sofia, Patrícia, Cátia, Tiago, Raquel e associados.

E, finalmente, à Andreia Filipa Silva Ribeiro, por ter sido a companheira de todas as horas, ouvido de todas as queixas e ser voz de todas as críticas. Não teria conseguido sem ela e sem o seu apoio constante e a sua motivação interminável. Para além desta tese, todo este percurso não teria sido possível sem ela. Um enorme obrigado à minha cientista favorita.

Contents

1. Introduction

1.1. Moral Judgement and Decision-making……11

1.2. Psychological research on moral judgment and decision making….12

1.3. The social-intuitionist model of moral judgement……....16

1.4. Research goals and objectives…….....20

2. Method

2.1. Participants……22

2.2. Materials……....…….…...22

2.3 Procedure…….……...24

3. Results

3.1. Manipulation Check…….....27

3.2. Control……..……..……...27

3.3. Agreement..…….…….…….....…..…..…….……28

3.4. Participant Answer……...……29

4. Discussion…….…..……..……..…….……..32

5. Final Remarks…….....……...39

6. References…….....……...40

7. Supplementary Materials…….…….…….…….…….……44

Abstract

Haidt's (2001) Social Intuitionist Approach to Moral Judgment not only claims for the prevalence of intuitive (deontological) over reasoned (utilitarian) judgments but also for the importance of social influence. However, the latter (i.e., social influence in moral judgment) has been seldom investigated. This is unfortunate since we rarely make moral judgments in a social vacuum but rather in interaction with others. Aiming to change this, we explored the impact of others' moral judgments on one's own judgments. By manipulating the perceived moral competence of the "other" (e.g., a Child vs. a Professor) and the moral orientation (deontological or utilitarian) of others' responses, we tested the prediction that participants (N=151) would show more agreement with high competence sources and sources that share their moral orientation; and give moral judgements accordingly. Results confirmed some of the predictions, namely the effect of the competence of the source on agreement and showed other effects of social context on participant's moral judgements.

Keywords: moral judgement, social context, social influence, decision-making.

Resumo

Esta tese surge no âmbito do vazio que existe na literatura no que diz respeito ao estudo do julgamento moral em contexto social. Haidt (2001) contraria a literatura existente e sugere um modelo completamente novo de julgamento moral, que contém uma componente social e cultural, mas que, porém, não explora desse ponto de vista. Nesta tese é abordado esse vazio, tentando explorar os mecanismos que estão por detrás do julgamento moral em contexto social, o seu funcionamento e factores importantes.

O julgamento moral é um tipo de julgamento complexo que envolve perspectivas pessoais de moralidade. Um exemplo de julgamento moral pode ser visto, por exemplo, em dilemas morais como o *Trolley Problem* (Thomson, 1986). Neste tipo de problemas, é apresentado um cenário hipotético que requere que uma decisão seja tomada, entre sacrificar uma ou poucas pessoas para salvar muitas, ou não. Há várias teorias que permitem explicar de que forma as pessoas resolvem este tipo de situações, mas, as mais comuns na anterior investigação em psicologia referem-se às perspectivas Utilitárias e Deontológicas. Estas duas perspectivas diferem totalmente: a perspectiva utilitária diz que a moralidade de uma ação é determinada pelas suas consequências, ou seja, causar dano a um indivíduo pode ser aceitável se salvar muitas outras pessoas. Por outro lado, a perspectiva deontológica afirma que a moralidade de uma ação depende exclusivamente da natureza da ação em si; causar dano a alguém é errado, não

interessa a quantidade de pessoas que pode salvar. Apesar de ser um pouco simplista, esta abordagem é a mais utilizada na literatura de forma a enquadrar os julgamentos morais dos indivíduos.

Antes de Haidt, a grande maioria da pesquisa em julgamento moral abordava o mesmo de um ponto de vista estritamente racional, vendo o processo como deliberado e dependente de funções cognitivas especificas. De facto, a investigação anterior focou-se na forma como, através da maturação cerebral, da experiência e da tomada de perspectiva, os indivíduos desenvolvem a sua moralidade (Piaget, 1965; Kohlberg, 1969). Posteriormente, novas investigações vieram mudar a forma como o julgamento moral é visto, causando uma mudança de paradigma que começou a olhar para os processos intuitivos e para as emoções como forma de explicar, em parte, o julgamento moral.

Recentemente, os psicólogos começaram a olhar para o julgamento moral mais como um conflito entre respostas emocionais e racionais. Nesta sequência, Haidt (2001) propôs um novo modelo de julgamento moral: o modelo socio-intuicionista. Este modelo sugere que a maioria dos julgamentos morais resulta de processos intuitivos que são causados pelas características emocionais dos dilemas, que funcionam como que se empurrassem os indivíduos para uma decisão instantânea de bom ou mau, de aprovação ou desaprovação de certos cenários ou ações. Este modelo aparece em oposição aos modelos racionalistas sugerindo uma forte componente emocional no julgamento moral; é intuicionista, visto que se foca no papel das emoções no julgamento moral e olhando para deliberação como um mecanismo secundário; é social, pois olha para os julgamentos morais como definidos pelo contexto social em que ocorrem. No entanto, esta perspectiva social foi sendo abandonada e, ao invés disso, a investigação evoluiu noutra direção, dando origens aos modelos dualistas de julgamento moral (Greene & Haidt, 2002).

Estes modelos dualistas de julgamento moral concluem que há uma forte componente emocional e intuitiva no julgamento moral, mas que processos de deliberação são também importantes. Baseando-se nos modelos dualistas de processamento cognitivo, o modelo dualista de julgamento moral faz uma ligação entre os dois tipos de processos (T1 – intuitivos, T2, deliberados) e as duas perspectivas morais (deontológica e utilitária): ligando julgamentos deontológicos a processos rápidos e intuitivos, e ligando julgamentos utilitários a processos lentos e deliberados. Este modelo dualista foi, no entanto, criticado por diversos outros autores (e.g. Bago et al., 2019; Mata et al., 2020).

O modelo de Haidt (2001) sugere que, num contexto social, haverá efeitos de influência, sendo que os julgamentos de outros poderão influenciar o julgamento dos indivíduos. Nesta linha,

teorias de Elaboração de Conflito (Mugny et al. 1995) afirmam que exposição a julgamentos de outros gera um conflito entre informação nova e informação prévia e que, nesta situação, os indivíduos tendem a aceitar a resposta da fonte mais competente ou da maioria. Ainda sobre a influência de outros no julgamento moral individual, Sperber el al. (2010) sugerem que as pessoas possuem mecanismos de Vigilância Epistémica que lhes permitem evitar ser induzidos em erros por outras fontes de informação, analisando a fidedignidade, a benevolência e a coerência da fonte e da informação, assegurando desta forma que a informação é vantajosa para o indivíduo.

Considerando esta Teoria da Elaboração do Conflito e os mecanismos de Vigilância Epistémica, é sugerido que quando a resposta de uma fonte a um julgamento moral é apresentada, a sua competência e a sua orientação moral vão definir a forma como os indivíduos avaliam a validade da sua resposta e, subsequentemente, resolver o conflito entre o seu próprio julgamento e esta nova informação, quando confrontados com um julgamento moral.

Há, portanto, dois pontos importantes que têm sido, de certa forma, ignorados na literatura existente, relativamente ao julgamento moral. Por um lado, 1) o tipo de resposta do outro pode influenciar a forma como um individuo percebe e responde a problemas morais. A orientação moral e extremidade desta resposta podem causar um conflito ou, pelo contrário, reforçar a resposta inicial do indivíduo. Por outro lado, 2) a percepção da competência moral de outros pode influenciar a forma como os indivíduos veem as suas respostas e, consequentemente, a forma como responde a problemas morais.

Devido a isto, nesta tese prevê-se que: (i) quando um julgamento moral de uma fonte mais competente é apresentado, os indivíduos devem concordar mais (e a sua resposta devia ser mais influenciada) com a resposta dada pela fonte mais competente do que com a fonte menos competente; (ii) os indivíduos devem concordar mais (e a sua resposta devia ser mais influenciada) com a fonte que partilha a sua orientação moral do que com a fonte que apresenta uma orientação moral oposta; (iii) o efeito descrito em (ii) deverá ser mais forte para a fonte mais competente comparando com a fonte menos competente, pois é esperado que os indivíduos avaliem a última como menos válida.

Para testar estas hipóteses, foi proposto um paradigma experimental que consiste na apresentação de dilemas morais a indivíduos. Antes de poderem responder, são apresentadas as respostas de duas fontes fictícias desenhadas para simular uma fonte social competente (Carlos, de 41 anos, professor de Filosofia) e uma fonte incompetente (Diogo, de 9 anos, estudante) que dão respostas extremas Utilitárias ou Deontológicas. Este paradigma tinha como

objectivo testar a hipótese da influência social ao confrontar os participantes com fontes de diferente competência e orientação moral.

No início da experiência os participantes responderam a três dilemas de controlo, de forma a estabelecer a sua *baseline*, relativamente a orientação moral. Posteriormente, viram um cenário fictício em que uma pessoa desconhecida cometia uma ação imoral. Eram apresentadas as respostas das duas fontes (Carlos e Diogo), ambos a reprovarem o comportamento e dando justificações diferentes. Depois, os participantes respondiam a quatro questões desenhadas para determinar o quanto achavam ser mais ou menos competentes do que ambas as fontes e a população no geral. Depois, eram aleatoriamente distribuídos por uma de quatro condições (Criança deontológica, criança utilitária, adulto deontológico, adulto utilitário). Era mostrado aos participantes a fonte com quem tinham sido emparelhados, mas não a sua orientação moral. Depois, os participantes viram seis dilemas morais, sendo-lhes apresentada a resposta da fonte ao dilema. Era-lhes pedido depois que dissessem o quanto concordavam com a resposta dada pela fonte e, por último, que dessem a sua própria resposta ao dilema.

Os resultados mostraram que os participantes tendem a concordar mais com a fonte competente do que com a fonte incompetente, e geralmente têm tendência a concordar mais com a fonte utilitária do que deontológica. No que diz respeito ao próprio julgamento dos participantes, participantes utilitários mostraram uma tendência para dar respostas mais utilitárias quando expostos a fontes utilitárias; no entanto, este efeito não se verificou em participantes deontológicos. Algumas das predições não se verificaram e surgiram outros efeitos que não haviam sido previstos. Ambos estes postos são discutidos neste trabalho.

Palavras-chave: julgamento moral, contexto social, influência social, tomada de decisão.

1. Introduction

The idea for this thesis arises from the relative void in the literature concerning the study of moral judgement and decision making in a social context. More specifically, although moral judgment is a widely investigated topic in psychology, little is known concerning how one's moral judgement is influenced by the judgments of others.

Haidt (2001) approached the longstanding discussion concerning lay people moral judgement by proposing the social-intuitionist model that, contrary to the prior research stream that defended the central role of rational reasoning as the main determinant of moral judgment (e.g., Kohlberg, 1969; 1971; 1976; Piaget, 1965), not only calls for the prevalence of emotions and intuition (above reasoning) in moral judgments, but also argues for the importance of social and cultural influences on lay people moral judgements.

However, in spite of addressing the importance of the social and cultural mechanisms towards moral judgement, Haidt's (2001) model and the research that followed failed to explore further the impact that the actual and proximate social context (i.e., the moral judgment of others) may have on individual moral judgements. This shall be the main focus of the present thesis.

In our daily life we may sometimes make moral judgement in social isolation but, more often than not, we have access to other people's judgements when we face moral dilemmas. Here I would like to begin exploring the extent to which others' moral judgements influence one's own judgements.

1.1. Moral judgement and decision-making

A trolley is barrelling down the railway tracks. Ahead, in the tracks, five people are tied and unable to move. The train is going straight at them and there is no way to warn them. However, a railway worker may pull a lever and change the direction of the trolley to another track, where one person stands. If the worker pulls the lever, the one person will die, but the other five will live.

This is a moral dilemma known as the Trolley Problem (Foot, 1967; Thomson, 1986), a hypothetical scenario in which a decision must be made, whether or not to sacrifice one or few to save many, and that is determined by the moral perspective of the actor.

There are a plethora of moral theories and approaches to account for people's responses to dilemmas such as this. The most frequently referred on the extant psychological research on moral judgment are the consequentialist approach and, in particular, the utilitarian perspective, that is often opposed to the deontological perspective.

The perspective of utilitarianism says that the morality of an action is determined by its consequences (Mill & Bentham, 1987). Specifically, utilitarianism follows the rule that the greater-good should always be pursued. Therefore, causing harm to an individual can be acceptable if it prevents harm to a greater number of people. Going back to the trolley problem, from a utilitarian point of view, the railway worker should pull the lever, killing one person but saving many more, maximizing the utility of his choice.

On the other hand, the perspective of deontology (Kant, 1785-1995) says that the morality of an action depends on the intrinsic nature of the action and not on its consequences. Human life is not quantifiable, so harming someone is considered wrong regardless of how many people it may save. From the deontological point of view the railway worker should not pull the lever, because sacrificing a human being is inherently wrong, regardless of the number of people such action may save.

A cursory review of the philosophical discussion on morality would be enough to understand that the abovementioned descriptions of both theoretical perspectives are a bit simplistic (Freeman, 1994; Kahane & Shackel, 2010). However, they correspond to the dominant views endorsed by current psychological research (e.g., Conway & Gawronski, 2013; Bago & De Neys, 2019; Greene & Haidt, 2002; Haidt, 2001) on the field of moral judgement, which has

relied on the use of moral dilemmas such as the trolley problem as a simple and effective way to experimentally study moral judgement and decision making. The present work also adopts these views and continues this experimental approach.

1.2. Psychological research on moral judgment and decision making

For a long time, the research on moral judgement remained dominated by theories of human development that gave an important role to reasoning and higher cognitive functions, first in children (Piaget, 1965), and later by analysing how morality works in adults (Kohlberg, 1969; 1976). Higher levels of moral development wouldn't arise simply from brain maturation, but it would imply experience, gained by taking roles and analysing moral issues through different perspectives (Kohlberg, 1969, 1976). This would improve the complexity and quality of one's moral reasoning and, consequently, moral judgement.

However, more recent theories and breakthroughs in psychology caused a shift from this rationalist approach to moral judgment. Research with primates (Flack & de Waal, 2000) suggested that emotions (e.g., shame, sympathy) might be the basis for moral judgement, making individuals care about each other, cooperate, and follow rules. Other research focused on the role of intuition and emotion (Damasio, 1994), contributing to a shift on the psychological perspectives on moral judgement. Specifically, advances in neuroimaging methods allowed researchers to begin to understand the neural bases of moral judgements (Dolan, 1999), through the analysis of decision-making and moral behaviour in patients with specific neurological lesions (Damasio, 1994; Damasio et al., 1994, Bechara et al., 1996). These neuroimaging studies present important findings concerning the basis of morality and moral judgement, by showing a link between emotion and decision-making.

More recently, both psychologists and philosophers have started looking at moral judgement more as a conflict between emotional and reasoned responses to moral problems.

Jonathan Haidt built on this research, combining it with studies about automaticity of socio-cognitive processes and behaviours (e.g., Bargh & Chartrand, 1999), to propose a novel model of moral judgement: the social-intuitionist model (Haidt, 2001). This model suggests that most moral judgements result from intuitive processes, that are triggered by the emotional laden

features of the moral dilemmas and that work as if they “compelled” us to make instantaneous decisions of good or bad, approving or disapproving certain moral conducts.

The social-intuitionist model is in opposition to the previous rationalist models (e.g., Kohlberg, 1969), that viewed moral judgment as stemming from deliberate reasoning, mostly dependent on experience and deliberate analysis of a moral problem, changing its scope towards the importance of emotion and context. Haidt’s initial research focused heavily on emotion; specifically, research involving “disgust” suggested that "cultural norms and culturally shaped emotions have a substantial impact on the domain of morality and the process of moral judgement” (Haidt, Koller & Dias, 1993). He and collaborators showed that people often get “morally dumbfounded” when they give a deontological answer to moral dilemmas, suggesting that deontological responses may not necessarily be Kantian responses (in the sense that they follow the strict rule to do no harm at any costs), but emotion-based responses instead.

Indeed, the social-intuitionist model is “intuitionist” as it focuses on the role of emotion on moral judgements and decision-making, regarding reason and rationality as a secondary mechanism, with little to no function. The model is “social” since it assumes that moral judgements are shaped by our social environments. Specifically, the social influence that others moral judgments play on the way we resolve moral problems. However, this social dimension was mostly disregarded by research on moral judgment. Instead, Haidt’s work evolved into other directions (Haidt, 2002; 2003), and eventually evolved into the emergence of dual models of moral judgement (Greene & Haidt, 2002).

Dual models of moral judgement rely on recent studies that investigated the cognitive processes underlying moral judgement and the role that emotions play in these judgments, often resorting to cognitive neuroscience. These studies suggest that emotion and intuition play a larger and more important role in moral judgement and decision-making than deliberate reasoning; nonetheless, deliberate reasoning still plays a significant role (Greene & Haidt, 2002). Accordingly, fMRI data suggests that both emotional and cognitive processes play a role on moral judgement, sometimes competing against each other in order to solve a difficult personal moral dilemma where personal moral values are being violated (Greene, Nystrom, Engell, Darley & Cohen, 2004).

More generally, dual-process perspectives on moral judgment have been influenced by more dual-process models of higher cognition (Evans & Stanovich, 2013; Kahneman, 2011), which

describe human cognition as a involving two types of psychological processes that act in combination: fast, largely autonomous and intuitive processes (T1); and slow, intentional and more deliberate (involving working memory) processes (T2). Dual-process models of thinking describe decision-making as the result of the interaction between the two types of processes. Specifically, the dual-process model of moral judgement advanced by Joshua Greene (Greene et al., 2008; 2013; Greene & Haidt, 2002) is a default-interventionist model that connects utilitarian judgments with the slow and deliberate T2 and the deontological judgments with the fast and intuitive T1. This model proposes that giving a utilitarian response to a moral dilemma demands that the T2 is activated, in order to override an intuitive T1 response to do no harm (Greene et al., 2008).

Behavioural studies provided further support to this theory. However, the studies that usually investigate moral judgement and its functioning rely heavily on answers to moral dilemmas. They usually design these dilemmas in such a way that the deontological (intuitive and emotional laden) and utilitarian judgments (deliberate and consequentialist) correspond to opposite responses in dichotomic choice paradigms (e.g. Bartels, 2008; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008). It is thus hard to understand if the two judgment processes are dependent on each other or if they contribute separately to open responses. Conway and Gawronski (2013) used the Process Dissociation Procedure (Jacoby, 1991), to address this ambiguity of "whether the obtained effects are due to differences in either deontological or utilitarian inclinations, or both" (Conway et al., 2013), by comparing responses to congruent trials (where both inclinations lead to the same response) and to incongruent trials (where emotional laden deontological inclinations and more deliberate utilitarian inclinations lead to a different response) and mathematically quantifying the relative influence of each process. This allowed them to show that both inclinations (deontological and utilitarian) provide independent contributions, resolving most of this ambiguity and further supporting Greene's (2008) dual model.

However, this dual process model to moral reasoning and moral judgement has been criticized as having some limitations. For instance, more recent research shows that utilitarian responses can be effortless and fast (Trémolière & Bonnefon, 2014), not requiring the amount of cognitive resources and time initially suggested. Also, the idea that people must activate deliberate thinking (T2) in order to override an intuitive decision to do no harm to others has been criticized by Białek & De Neys (2017), who showed that T1 can cue both a deontological and

utilitarian intuition. Another problem with this model is that the actual processing specifications of the T1 and T2, concerning moral judgements, are not clear (Bago et al., 2019; Mata, Vega, Ferreira & Vaz, 2020). Additionally, Mata et al. (2020) showed that the extremity of the answer also plays an important role concerning moral judgements. Specifically, their study found that more intuitive answers, with a higher FOR (Feelings of Rightness) are more extreme, regardless of those answers being deontological or utilitarian. This further suggests that T1 is not exclusively linked to deontological judgements and shows that the extremity of the answers plays an important role in moral judgement.

Finally, in spite of addressing important issues concerning the mechanics of moral judgement, the research literature (including research on dual-process models of moral reasoning), still lacks one important point: how does the social context, more specifically having access to others' judgments, influence individual moral judgement?

1.3. The social-intuitionist model of moral judgement

Haidt's (2001) social-intuitionist model may still be a promising way to approach this issue. The model assumes that, if a moral judgement occurs in a social context, there will be social influence effects, in which other people's moral judgements will influence our own judgement, even in the absence of any justification or rational arguments. Furthermore, it is also assumed that these social influence effects should be translated into some form of social conformity, although the model is mostly mute on how social influence and social conformity are expected to affect individuals' reasoning and judgment.

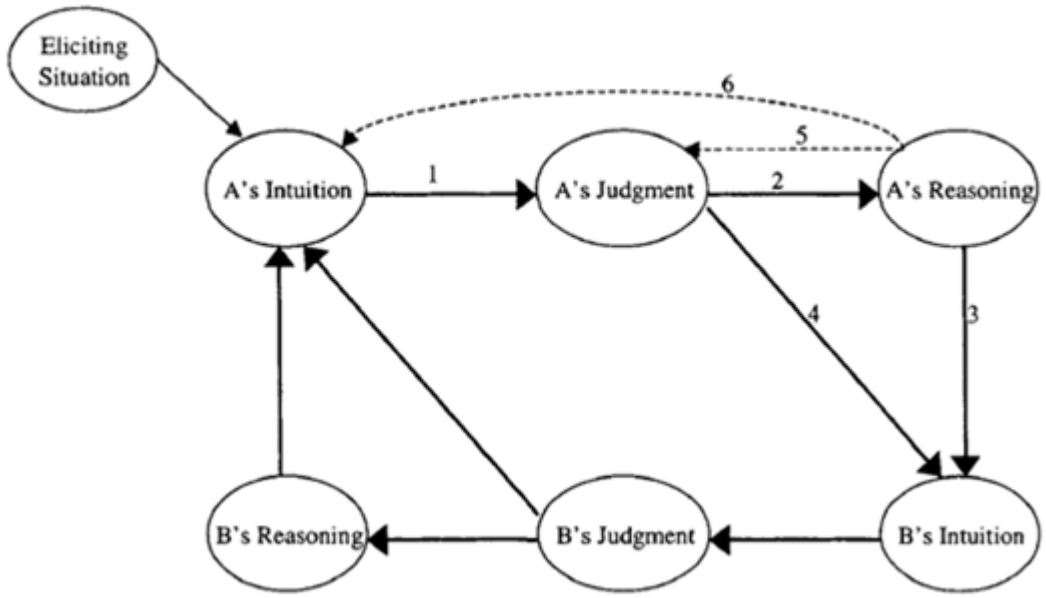

Figure 1. Social-intuitionist model of moral judgement. The numbered links (for Person A) are (1) the intuitive judgment link, (2) the post hoc reasoning link, (3) the reasoned persuasion link, and (4) the social persuasion link. Two additional links are hypothesized to occur less frequently: (5) the reasoned judgment link and (6) the private reflection link. Adapted from Haidt (2001, p. 815).

The influence of others on one's moral judgment may be conceptualized as the resolution of a potential conflict (or at least disparity) between judgments. The conflict arises because individuals do not simply accept new judgments uncritically, especially when they already hold their own.

In fact, according to the conflict elaboration theory of social influence (Mugny, Butera, Sanchez-Mazas, & Perez, 1995) exposing an individual to others' judgments that are different from the individual own judgment, generates a conflict that leads to questioning the validity of the new, divergent judgments. In this situation, the individual tends to accept the responses that originate from the highest-status source (e.g., the majority, a high competence source) as more

valid. On the other hand, when the same responses are provided by a low-status source (e.g., a minority, a low competence source), the individual tends to disregard it (Butera, Legrenzi, Mugny, & Perez, 1992; Butera, Mugny, Legrenzi, & Perez, 1996).

In the same vein, classic dual-process persuasion theories focus on two major aspects: the source of the information and the message (or the contents of the information itself), often composed of arguments against or for a given position (Chaiken, 1980; Petty & Cacioppo, 1986).

Here, our focus is on the conflict that may arise between the individual and the source responses and on the effects that may follow. The same conflict will have different effects depending on 1) the credibility source is (e.g., high versus low competence sources) and 2) the type of judgment task (Mugny et al.,1995; which will be characterized later on).

Another relevant theoretical framework to understand the influence of others on one's own moral judgments, is the work Daniel Sperber and collaborators. According to them, the main function of reasoning is argumentative (to persuade) and hence is highly dependent on human communication (Mercier & Sperber, 2011). As such, rather than simply accepting judgments communicated by others, people have mechanisms of epistemic vigilance that allow them to avoid being (intentionally or unintentionally) misled by other social sources of information (Mercier & Sperber, 2011, 2017; Sperber, Clément, Heintz, Mascaro, Mercier, Origgi & Wilson, 2010). One of these mechanisms - trust calibration – involves being vigilant to the trustworthiness of the source of the communication, with respect to how benevolent or competent the other is. Another mechanism – coherence checking – involves being attentive to the coherence between the judgment communicated by others and one's previous knowledge and beliefs (Sperber et al., 2010). In sum, epistemic vigilance corresponds to a suit of cognitive mechanisms that safeguards misleading information from being accepted and ensures that communication remains advantageous to the individual.

All these approaches, dual-process theories of persuasion, the conflict elaboration theory and the epistemic vigilance mechanism, allow us to complement Haidt's social-intuitionist model and to address one of the points hypothesised in his model: How may other's judgments influence the way we perceive and resolve individual moral dilemmas?

Specifically, can the other's moral inclination to answer in a more deontological or utilitarian way change our own answer to a moral dilemma? Or, on the contrary, can it even reinforce our

initial judgment? According to the theoretical approaches briefly reviewed, this may depend on whether the other is perceived to be more or less morally competent (or incompetent).

In fact, any form of social influence caused by the moral judgements of others is likely to depend on the way we perceive the other. We can see the other as more or less morally sophisticated (more or less driven by their emotions) than we are – when dealing with emotion laden moral problems that demand controversial or moral code breaking decisions. Such social perception of others' morality may influence the way we think about moral dilemmas.

Here I explore these issues using different moral dilemmas that share the same deep structure of the Trolley Problem but vary in their superficial structure. In these dilemmas, two groups (each with different numbers of people) are in direct danger of dying or getting seriously hurt. An actor must decide whether to let the larger group die or actively sacrifice a smaller amount of people in order to save the larger group. This problem constitutes a violation of our moral code, forcing an individual to make a difficult and controversial decision that always results in the loss of human lives. It is assumed that how people resolve the conflict depends on their moral orientation. As such, these dilemmas allow us to explore the impact of the social context (i.e., the moral judgment of others) in the participants' own judgments.

As aforementioned, how social influence processes shape the conflict between the source's and own response depends on the nature of the task at hand (Mugny et al.,1995). It is thus important to consider the type of task the dilemmas are.

According to Mugny et al. (1995) such classification is based on two dimensions: whether the task allows for only one right answer or not; and whether the tasks' responses socially anchor the targets or not. Four distinct tasks follow from the cross-over of these two dimensions: TAP (Tasks APtitudes), TOP (Tasks OPinions), TONA (Tasks Objective, and UnAmbiguous) and TANI (TAsks Non-Implicating).

TAP imply that one answer must be right (or righter than others) and is a socially anchoring task because people will be concerned about the correctness of their judgements and their self-image. Social influence in TAP often translates into informational dependence (Deutsch & Gerard, 1955). The individual tends to perceive the source's response as more valid when the source is perceived as competent, leading her to imitate it. When the source is perceived as incompetent, there is a conflict of incompetences, that eventually leads the individual to

distance himself from the source's response and to look for a (another) response (focusing on the task information).

TOP are socially anchoring tasks where there is not necessarily a right or wrong answer. Because of that, normative dependence tends to prevail (Mugny & Doise, 1978). So, for instance, conflict with an ingroup member leads to conformity, while conflict with an outgroup member leads to the reinforcement of the original response or an increase in the difference.

TONA are non-socially anchoring tasks that must have a correct answer or an answer more correct than others (e.g., Solomon Asch conformity paradigm; Asch, 1956). In this case conflict elaboration will be mainly determined by the existence or absence of unanimity, resulting in conflict any time someone does not agree, simply because that unanimity is always expected regardless of expertise.

Lastly, TANI are non-socially anchoring tasks where there is not necessarily a correct answer. Because of this, people's judgements present no threat to self-image since errors are irrelevant. This results in a lack of expectation to reach a consensus and, therefore, lack of conflict.

Although an argument can be made that there is no right answer to moral dilemmas such as the trolley problem (TOP) and that different (even opposite) moral judgment may be equally justifiable, moral issues are usually perceived as having an answer that is more valid than the rest (TAP) at least in the Piagetian sense that judgment and decisions stemming from more sophisticated moral reasoning are more valid than the ones stemming from simpler and more biased forms of reasoning (Kohlberg, 1971; Piaget, 1965). As such, social influence (e.g., informational dependence) effects typically observed in TOP and to a certain extent in TAP (e.g., normative dependence) are likely to be relevant to account for how others' responses may influence one's responses to moral dilemmas.

1.4. Research goals and objectives

There are two main points that have, so far, been in one way or the other left out in the existing literature and that represent the social core of the moral judgement and decision-making research. On the one hand, 1) the other's type of response might influence the way an individual thinks and answers to moral dilemmas. Such influence is likely to depend on the extremity and moral orientation (more deontological or utilitarian) of the other's answer in relation to the individual's answer. Specifically, the other's answers may be more or less extreme than the individual's answer and may be in conflict (e.g., one is utilitarian and the other is deontological) or aligned (e.g., both answers are utilitarian, or both are deontological) with the individual's answer. This might cause a conflict or, on the contrary, reinforce the existing moral orientation. Furthermore, 2) our perception of the other's moral competence might also influence the way the individual sees the other's response and, consequently, the way the individual answers. Specifically, when others are perceived to be more morally competent their responses to moral dilemmas may lead to more conformity and less reactance than when the others are perceived as less morally competent. On this issue, both the Conflict Elaboration Theory (and dual-process persuasion theories) and the Epistemic Vigilance mechanisms suggest that, when the answer of a source to a moral judgement is presented, that source's moral competence (higher or lower competence) and its moral orientation will dictate how individuals evaluate the validity of such judgement, and subsequently resolve the conflict between their own with judgement and this new information.

These two points may interact, changing or reinforcing one's moral judgement depending on the way others are perceived in terms of competence and on the alignment (or conflict) between the individual and the other's moral orientation (more deontological or utilitarian). Or they may have no effect at all, which would indicate that the social context has little to no effect on the way individuals make moral judgements.

Specifically, in this research it is hypothesized that:

(i) when a moral judgement from a higher competence source is presented, individuals should agree more (and their answer should be more influenced) with the answer given by the high competence source than when the same answer is given by a lower competence source. This is because the conflict between new and old information is expected to result in the lower competence source to be disregarded more easily.

(ii) An individual should agree more (and their answers be more influenced) with the source that shares the individual's moral orientation than with the source that presents an opposing moral orientation. This is because the conflict that arises from opposing moral orientations should be absent when the orientations are equal, resulting in an easier acceptance of new information.

(iii) The effect described in (ii) will be stronger for the high competence source compared to the low competence source, because individuals are expected to perceive the latter as less valid than the former.

To test these hypotheses, I proposed an experimental paradigm that consists in confronting the participant with the answers of a source (i.e., of another alleged participant who participated in a prior session of the same study) to several moral dilemmas similar to the Trolley Problem. For each dilemma a more competent (a 41 years old philosophy professor) or less competent (a 9 years old student child) source in moral issues could give extreme utilitarian or deontological answers. The participant is then invited to express in a rating-scale how much he agrees/disagrees with the decision made by the source, and afterwards gives his own response to each moral dilemma. The rationale is that by letting participants know what the source answered to the dilemma before they can answer themselves, social influence may arise and change or reinforce their intended answer, depending on the competence of the source and degree of conflict between the source's responses and the participants own moral judgment. A measure of Bias Blindspot (Pronin, Lin & Ross, 2002) would be introduced in order to understand if the participants position themselves to be more morally competent (or less emotionally biased in moral issues) or not than the sources, to understand if the manipulation of the source's perceived moral competence worked.

In summary, in this study I intended to understand 1) how other people's answers to moral dilemmas may influence individual moral judgments; and 2) how the moral inclination and perceived moral competence of others (in relation to the moral orientation and competence of the individual) may shape the individual moral judgments.

2. Method

2.1. Participants. One hundred and sixty participants (M_{age}=26.8, *SD*=5.8) were recruited via the web-based platform *Prolific*. Data from one participant who did not comply with the instructions was discarded, leaving a total of 159 participants.

2.2. Materials. *Moral Reasoning Problems*. In total, nine moral reasoning problems were presented, where three were control problems and the following six were actual testing dilemmas. The content of the problems was based on classical moral dilemmas from the literature (e.g., Cushman, Young, & Hauser, 2006; De Neys, 2019; Royzman & Baron, 2002). The same structure was used for every moral dilemma, requiring subjects to decide whether they would sacrifice the lives of one out of two groups of characters in the scenarios or do nothing at all. In order to minimize possible identification issues (Bago et al., 2019) or personal conflicts with the presented scenarios, the following rules were applied to every problem: (1) the scenario never involved the participant or their family, (2) the difference between the number of characters on the two groups was kept to a small number (between 1 and 15), (3) all characters were adults and (4) all sacrifices mentioned in the scenarios resulted in the death of the characters.

The problems were presented and then answered in three parts. First, the actual moral reasoning problem was presented to the participant and they were asked to click a button when they were finished reading it. Then, participants were shown on the next screen the answer of one of the sources to the presented problem, consisting of their assessment of the action taken in the scenario as acceptable or unacceptable. They were then asked to answer how much they agreed with the answer the source gave (on a scale of 1- Totally Disagree to 9- Completely Agree) and, finally, they were asked on another screen to give their response to the moral dilemma, answering (on a scale of 1- Definitely Not to 9- Definitely Yes) if the actor depicted on the scenario should go ahead and perform the action.

Here is an example of one of the moral reasoning problems:

“There is a flood near a village with 10 inhabitants. If nothing is done, the whole village will be destroyed and everyone will die. There is not enough time to move them. Fortunately, an engineer has the power to close a floodgate, diverting the flow of the river towards a lone house, where 2 people live. If he closes the floodgates, the 10 villagers will live, but the 2 people inside the lone house will die. Should the engineer close the floodgates?”

“In the previous problem, X (the source) said that it was completely acceptable for the engineer to close the floodgates. Do you agree with this answer?”

“Now, give your own response to the problem. Should the engineer close the floodgates?”

In order to maintain attention and to ensure that people were truthfully answering the dilemmas on the manipulation phase, three traditional “conflict” dilemmas and three “no-conflict” dilemmas were used. In the conflict dilemmas, like the one mentioned above, participants were asked whether they would sacrifice a small number of people to save a larger group of people. In the “no-conflict” dilemmas, participants were asked whether they would sacrifice a large group of people to save a small number of people (e.g., Bago & De Neys, 2019).

Here is an example of one of the “no-conflict” moral reasoning problems:

“Due to a malfunction on a power plant, a leak of a toxic gas threatens to enter a small room where to scientists are working. If nothing is done, they will die from exposure to the toxic gas. The only way to save these 2 workers is for the chief-engineer to press a button that vents the toxic gas to the outside through a chimney. However, he realises that there are 8 construction workers repairing the chimney that will die if the toxic gas is vented through the chimney. Should the engineer vent the gas?”

These dilemmas were also presented with the same questions as the “conflict” dilemmas:

“In the previous problem, X (the source) said that it was completely unacceptable for the chief-engineer to vent the gas. Do you agree with this answer?”

“Now, give your own response to the problem. Should the chief-engineer vent the gas?”

Hence, on the “conflict” version the utilitarian response is to answer “yes” and the deontological response is to answer “no.” On the no-conflict problems, both utilitarian and deontological considerations cue a “no” answer. For example, a more utilitarian person should not, in a real-world situation, go out of their way in order to actively sacrifice 10 people in order to save 1. Also, I included the “no-conflict” versions to make the problems less predictable and avoid that participants would start to reason about the possible dilemma choice before presentation of the second part of the problem. For the same reason a filler item was also included in the middle of the experiment (i.e., after 4 test problems). In this filler problem

saving people did not involve any sacrifice (i.e., doing the action implied saving 6 and killing 0 characters).

All problems are presented in the supplementary materials section.

Source's Morality Problem (Thief Problem). In order to establish the sources' perceived morality and to reinforce the manipulation, a story about a fictional robbery was created that presented a random unknown person (John) doing a morally wrong action. The participants would then be asked to rate on a 9-point scale (1 - Completely Disagree; 9 - Completely Agree) the degree to which they agreed or disagreed with the action shown on the story. Here is the fictional robbery scenario:

"John is going down a street on a deserted Lisbon suburbia when he sees, a few meters ahead, an elderly woman who drops her wallet and goes insider her house. Inside the wallet is the elderly woman's retirement money that she just received, totalling 800€ in bills. John picks up the wallet and counts the money. Nobody sees him, there is no surveillance whatsoever and no kind of technology on the wallet allows its location to be known. John decides to keep the wallet and the money, and to leave that place as fast as possible."

Cognitive Reflection Test. All participants answered a Cognitive Reflection Test (Frederick, 2005), consisting of three questions, adapted from standard lists present in several research papers. Here is an example of one of the questions:

"Ana and Sofia are picking seashells on the beach. In total, they picked 12 seashells. Ana picked 10 more seashells than Sofia. How many seashells did Sofia pick?"

A score was then assigned to each participant, consisting of the number of correct answers to the three questions (0 - 3).

2.3. Procedure. Participants began by responding sequentially to 3 control moral dilemmas. On each trial they considered a different dilemma and made a judgment concerning the acceptability of the utilitarian course of action (1 – Completely Unacceptable, to 9 – Completely Acceptable). We avoided using a dichotomous scale in order to increase response variability, and to be able to differentiate between extreme and moderate responses tendencies. These initial dilemmas were used to individually establish the participants' moral orientation (more utilitarian or deontological).

They were then submitted to a Bias Blindspot measure, which consisted in four questions designed to determine if they evaluate themselves to be more competent (or less emotionally biased) than the sources. A short sentence stated that psychologists say that people's moral judgements tend to be biased by the emotions triggered by the situations they encounter. Afterwards, they were asked to rate on a scale (1- Nothing to 9 - A lot) if they believed themselves to show this tendency. Afterwards, they were asked to what degree they believed the Adult Source showed this tendency. Next, they were asked the same for the Child source and, finally, they were asked to which degree they believe that the general population displays this effect.

Next, participants were told that on this Study we were interested to know more about how people make moral judgments in a social context. For that they would be presented with other moral dilemmas and besides giving their own answer they were going to have access to the judgment of other individuals who had (allegedly) participated in the Study before. Two response sources (represented by different avatars) were created: Diogo, a 9-year-old student and Carlos, a 41-year-old Philosophy professor. Participants were also requested to answer the same biographical questions (first name, age and profession) and asked if they would allow their moral judgments to be used (i.e., to be presented to other participants) in future sessions of the study. For this they also selected the avatar (from a set of possibilities) that would represent them in future trials. Next, the Thief Problem appeared on the screen with instructions to read carefully and answer truthfully. Participants expressed their agreement with John's (the actor in the story) behaviour using a 9 point rating scale (from 1 – completely disagree, to 9 – completely agree) and gave a brief justification for their response. Next, they had access to the responses of the two sources created for the purpose of simulating a social environment: 9 years old Diogo and the 41 years old Carlos. Both strongly disagreed with John's behaviour (Diogo's response was "1" and Carlos response was "2" in the rating scale). Diogo justified his answer saying "Because the lady is very old and is going to be very sad", whereas Carlos justified his answer by saying "It's everyone's moral duty to act justly and to return the wallet and its contents to the owned, in a situation like this."

Participants were then told that they would be randomly paired with Diogo or Carlos for the remaining of the Study. After a short prompt, participants were randomly assigned to one of four conditions, which differed on the source (Child or Adult) of the responses and the type of moral response (extreme Utilitarian or extreme Deontological), resulting in the following

combinations: Deontological Child, Utilitarian Child, Deontological Adult and Utilitarian Adult. In the case of the Deontological moral response, the source would completely disagree with the action shown on the dilemma. In the case of the Utilitarian moral response, the source would completely agree with the action shown on the dilemma. The participants were shown the name of the source, but not the type of moral response.

Finally, participants were asked to carefully read the six moral dilemmas. In each trial (for each dilemma) they first had access to the source's response and expressed the degree to which they agreed with it (on the same 9-points rating scale, from 1 – Completely Disagree, to 9 – Completely Agree), and then gave their own answer to the dilemma by judging whether or not the main agent in the dilemma should follow the utilitarian course of action (save more by killing few). For this, they used a 9-points rating scale (from 1 – Definitely Not, to 9 – Definitely Yes).

Participants also completed the Cognitive Reflection Test (Frederick, 2005).

After the study, all participants were debriefed concerning the manipulation. The study received approval from the Faculty of Psychology of University of Lisbon's Ethics Committee.

3. Results

3.1. Manipulation Check. The average score of the responses to the Thief Problem was M_{tp}=1.28, indicating that participants clearly disagreed with John's action, classifying it as unacceptable. A measure of Bias Blindspot followed, where participants classified the child (M_c=7.35) as being more emotional and the adult (M_a=5.74) as being less emotional than themselves (M_s=6.31) when considering moral issues. Furthermore, participants also classified the general population (M_g=6.75) as more emotional than themselves. The three Bias Blindspot Scores were calculated computing the difference between how much the participants characterized themselves to be biased when compared to others (the child, the adult, and the general population). A one-way ANOVA was performed with the three BBS scores as factors (difference between participant and child, participant and adult, participant and general population), showed a significant effect, $F(2, 316) = 38.43$, $p < .001$. A planned contrast between the Bias Blindspot score of the Child (M_c=1.04) and the Adult (M_a=-.57) was also significant, $F=(1, 157) = 59.65$, $p < .001$, indicating that the participants perceived themselves to be more morally competent (less biased) than the Child in comparison to the Adult.

3.2. Control. In order to establish the initial moral orientation for the participants, the average score of the responses to the control dilemmas was calculated. Participants with an average score (≥ 5) were considered a utilitarian oriented, while participants with an average score below the midpoint (< 5) were considered deontological oriented. Results from the control dilemmas show that a utilitarian orientation was more frequent (N=118) than a deontological orientation (N=41), with an average response of M_u=5.57 for utilitarian participants and an average response of M_d=3.98 for deontological participants. Based on this classification, two groups of participants (Utilitarian and Deontological) were considered for the remaining analyses. Figure 2 shows the comparison between the average scores for the dilemmas on the Control phase, Agreement with the Other's Moral Judgement and Participant Moral Judgements, for the participants in every condition.

	Adult		Child	
	Deontological	Utilitarian	Deontological	Utilitarian
Control	5.27	5.52	5.6	5.58
Manipulation (Agreement)	4.47	7.59	3.51	7.01
Manipulation (Participant Answer)	5.78	7.40	6.47	7.08

Figure 2. - Average participant responses to the moral dilemmas on the control phase (before exposure to social context) agreement on the manipulation phase after exposure to social context (Score range 1- Totally Disagree to 9- Completely Agree) and participant answers on the manipulation phase (Score range 1- No, definetly to 9- Yes, definetly).

3.3. Agreement with the Other's moral judgment. A 2x2x2 ANOVA with Source (Child and Adult), Source Answer (Deontological and Utilitarian) and Self (Deontological and Utilitarian) as factors and Agreement as the dependent variable, showed two main effects and one interaction. A main effect of Source Answer $F(1, 151) = 100.99$, $p < .001$, such that participants showed more agreement with the Utilitarian Source Answer ($M_{us}=7.29$) than with the Deontological Source Answer ($M_{ds}=3.98$); and a main effect of Source $F(1, 151) = 9.40$, $p < .001$, such that participants showed more agreement with the Adult Source ($M_{as}=5.99$) than with the Child Source ($M_{cs}=5.24$), supporting our first hypothesis. There was also an interaction between Source Answer and Self, $F(1, 151) = 17.008$, $p < .001$, revealing that Utilitarian participants disagree with the Deontological Source ($M_{ds}=3.72$) and agree with the Utilitarian source ($M_{us}=7.39$) to a greater extent than Deontological participants ($M_{ds}=5.16$, $M_{us}=6.69$). There was no significant effect of the interaction between Source and Source Answer, contrary to what was predicted.

3.4. Participant moral judgments. A 2x2x2x2 ANOVA with Source (Child and Adult), Source Answer (Deontological and Utilitarian), and Self (Deontological and Utilitarian) as between factors, Judgement (Initial Judgement and final Judgement) as a within factor and participant answer as the dependent variable showed three main effects and three interactions.[1]

A main effect of Self, $F(1, 151) = 120.49$, $p < .001$, such that Utilitarian participants naturally showed more utilitarian answers (M_{su}=6.86) than Deontological participants (M_{sd}=5.67). A main effect of Source Answer $F(1, 151) = 12.53$, $p < .001$, such that participants showed a tendency towards more utilitarian answers when the Utilitarian Source Answer was presented (M_{us}=7.24) than when the Deontological Source Answer was presented (M_{ds}=6.13). A main effect of Judgment, $F(1, 151) = 33.99$, p <.001, revealing that participants showed more Utilitarian answers on their Final Judgement (M_{sj}=6.35) than on their Initial Judgement (M_{ij}=5.6). Furthermore, an interaction between Source and Source Answer, $F(1, 151) = 4.43$, $p = .036$, revealed that participants showed a greater tendency towards more utilitarian (M_{uas}=6.48) and towards more deontological answers (M_{das}=5.48) when the other's Utilitarian and Deontological answers came from an Adult Source then when these answers came from a Child Source (M_{ucs}=6.11 M_{dcs}=5.85).

An interaction between Judgement and Self, $F(1, 151) = 24.88$, p <.001, such that Deontological participants answers became more Utilitarian in the Final Judgement (M_{sj}=5.7) compared to the Initial Judgement (M_{ij}=4.31), while Utilitarian participants showed virtually no change in their Final Judgement (M_{sj}=7.01) compared to their Initial Judgement (M_{ij}=6.9).

An interaction of Judgement and Source Answer, $F(1, 151) = 18.22$, p <.001, such that participants answers became more utilitarian in the Final Judgement (M_{sj}=6.94) compared to the Initial Judgement (M_{ij}=5.64) when the Source Answer was Utilitarian but show almost no change when the Source Answer was Deontological (M_{ij}=5.56; (M_{sj}=5.77).

Figures 3 and 4 present the global pattern of results for Utilitarian and Deontological participants, respectively.

[1] I repeated this ANOVA using as dependent measure only participants final judgment and including the factor Self or including initial judgment as a co-variable (ANCOVA). The main results converge in the same results pattern (see supplementary material, section A). This was done because the ANOVA here reported included participants initial judgments as one level of the Judgment factor and the same variable (initial judgments) was used to create the Self factor. This may create dependency between the factors.

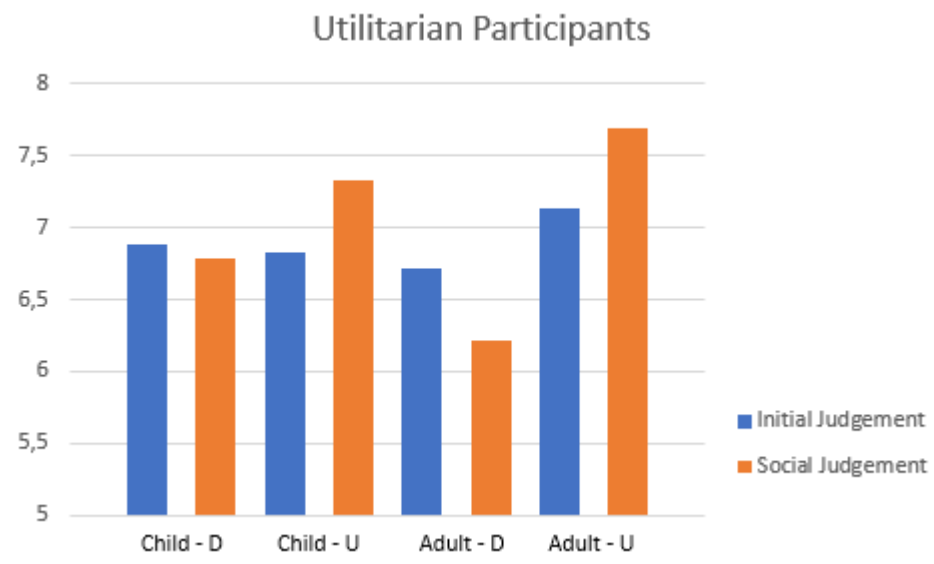

Figure 3. Initial and Final Judgements for Utilitarian participants (Self - utilitarian), separated by condition. (Source and Source Answer).

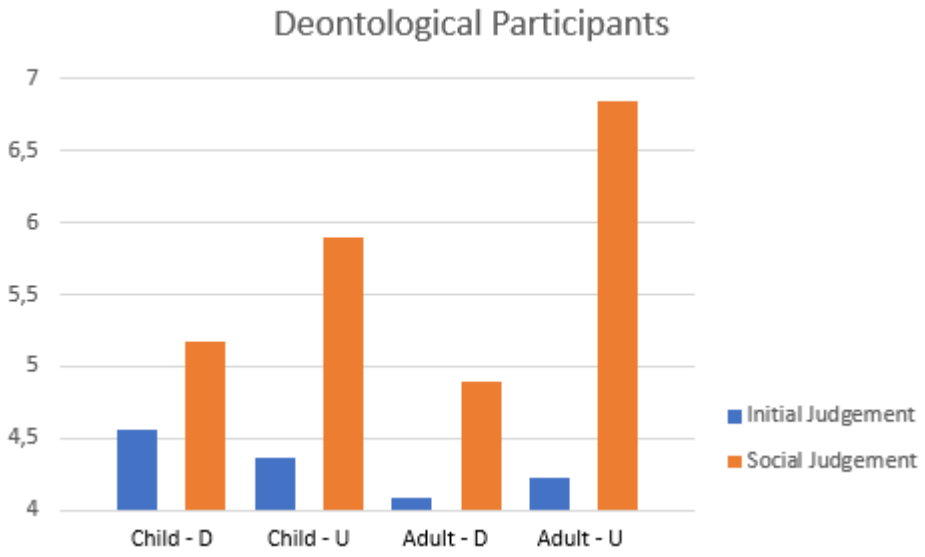

Figure 4. Initial and Final Judgements for Deontological participants (Self - deontological), separated by condition. (Source and Source Answer).

To further test the impact of the other's expertise and moral judgments on the participant's own judgments, planned contrasts were computed, separately for Utilitarian and Deontological participants, comparing Initial and Final Judgments when the other was: a) an Adult giving a Deontological response; b) and Adult giving an Utilitarian response; c) a child giving the deontological response; d) a child giving an Utilitarian response. For Utilitarian participants, there was a significant difference between Initial and Final Judgements when the Adult gave a Utilitarian response, $F(1, 151) = 4.01$, $p = .046$, indicating that the participants gave more

utilitarian answers in their Final Judgement than in their Initial Judgement; when the Child gave a Utilitarian response there was a significant difference, $F(1, 151) = 4.32$, $p = .039$, showing that participants gave more utilitarian Final Judgements; and finally a marginal effect for when the Adult gave a Deontological Answer, $F(1, 151) = 3.71$, $p = .055$ such that participants gave more deontological Final Judgements when this source was present. When the Child gave a Deontological Answer, the result was not significant, $F < 1$.

For the Deontological participants, there was a significant difference when the Child gave a Utilitarian response, $F(1, 151) = 8.79$, $p < .001$, showing more utilitarian Final Judgements compared to Initial Judgements; when the Adult gave a Utilitarian response a significant difference was identified, $F(1, 151) = 46.61$, $p < .001$, indicating that participants gave more utilitarian Final Judgements when this source was present; and a difference when the Adult gave a Deontological response, $F(1, 151) = 4.59$, $p < .03$, showing that participants also gave more utilitarian Final Judgements when this source was present. When the Child gave a Deontological response, however, the effect was not significant, $F(1, 151) = 1.53$, $p = .217$.

4. Discussion

This thesis focused on understanding the effects of social context on individuals' moral judgements. The study of this effect consisted on assessing how one's moral judgements affect the moral judgements of others and what factors moderate that effect. Going back to Haidt's model, others' moral judgments (even without arguments – i.e., without having access to others reasoning) have an effect on individual's judgments. Overall the results showed that this effect can be predicted by specific factors, as the influence of the judgements of others is dependent not only on their moral competence, but also on their moral orientation and the moral orientation of the participants.

The results seem to be in line with the first hypothesis, *(i) when a moral judgement from a higher competence source is presented, individuals should agree more (and their answer should be more influenced) with the answer given by the high competence source than when the same answer is given by a lower competence source.* Generally, a more trustworthy source (the Adult, a Philosophy Professor - perceived as more morally competent) led to more agreement than a less competent source (the Child, a 9-year-old student - perceived as less morally competent). Morality of the source (Deontological, Utilitarian) and competence (Philosophy professor, 9 years old child) interact, showing that for people's judgments, social influence is also greater for competent sources than for incompetent sources.

These results corroborate our prediction that source competence should influence the participants into agreeing with more competent sources to a larger degree than with the less competent sources.

Furthermore, this result seems to be in line with the theories presented in the introduction. Epistemic Vigilance mechanisms, in particular trust calibration, predicts that the individuals will use the other's competence in order to decide if they accept the information provided or not (Sperber et al., 2010). Dual process theories of persuasion (e.g., Petty & Cacioppo, 1986) also predict that the influence of the source depends on source credibility (although the present context is not exactly one of persuasion). According to the conflict elaboration theory, in the case of the conflict between the information originating from the individual and the information origination from a low or high competence, the individuals chose the high competence source as more valid (Mugny et al., 1995). On a related note, the way this conflict was resolved seems to indicate that individuals treated the moral dilemmas as TAP (Tasks APtitudes) since it

appears that, in this case, the source causes informational dependence, which is typically found in TAP tasks.

The second hypothesis was *(ii) an individual should agree more (and their answers be more influenced) with the source that shares the individual's moral orientation than with the source that presents an opposing moral orientation.* This prediction was not confirmed because the influence of others extreme Utilitarian and Deontological responses was not symmetrical, as predicted. Instead, Utilitarian sources had a greater influence than Deontological ones. In addition, participants tended to answer in a more Utilitarian fashion (from initial to final judgments) regardless of the source moral orientation.

Concerning the Agreement with the Other's Moral Judgements, there are some interesting results to be discussed. I expected that Utilitarian participants would agree more with the Utilitarian Source Answer and Deontological participants would agree more with Deontological Source Answer. The results from the interaction between Source Answer and Self show that the predicted effect is clearly present for Utilitarian participants (which agreed more with the Utilitarian source) but the same effect was not observable in the case of Deontological participants. Instead, Deontological participants also showed a slight increase in agreement, compared to their own moral orientation. This seems to be caused by the apparent greater impact that Utilitarian Source Answers have in participant's answers when compared to Deontological Source Answers. Contrary to the prediction that individuals would agree more with their own moral orientation, they instead seem to agree more with the Utilitarian orientation overall, attenuating the effect when Deontological Source Answers are present.

Concerning participants moral judgments, planned comparisons computed separately for Utilitarian and Deontological participants, show that the Utilitarian participants gave more Utilitarian responses in their Final Judgement in comparison to their Initial Judgement only when they were exposed to the Utilitarian Source Answers (both Adult and Child); indicating that these participants were influenced to answer towards the same orientation of the source, regardless of competence. This result seems to confirm the prediction that, due to the lack of conflict between different moral orientations, the individual's moral views are otherwise reinforced.

However, when we consider the Deontological participants, the results don't align with the prediction that they would give more deontological answers when a deontological source was

present. In fact, Deontological participants also showed a significant difference between Initial and Final Judgements, giving more utilitarian answers when exposed to all combinations of sources, except for the Deontological Child. This result is quite interesting because, as aforementioned, it suggests that Deontological participants not only tend to generally give more utilitarian answers regardless of the social source's moral orientation, but that the same effect is smaller when the source is perceived as less morally competent and shares the participant's moral orientation. Moreover, the fact that even for the Deontological participant's judgements, the Utilitarian source seems to have a greater impact than the Deontological source seems to be compatible with the higher agreement that these Deontological participants showed towards Utilitarian sources.

These results further expand on the existing discussion on morality and whether deontological individuals tend to become more utilitarian merely as the result of being exposed to more moral dilemmas. On the other hand, it might be the case that participants in general (and Deontological participants in particular) engage in more deliberation when another source is present, regardless of the source's moral orientation. Future studies using the same experimental paradigm of the current study could include a control condition (where participants make both the initial and final judgment without having access to others judgments).

Concerning the hypothesis that *the effect described in (ii) will be stronger for the high competence source compared to the low competence source, because individuals are expected to perceive the latter as less valid than the former*, globally the results do not support our prediction, as the expected interaction between Source, Source Answer and Judgement was not present.

Specifically, for the Agreement with Other's Moral Judgements, it was expected that there would be a significant interaction between the Source Answer (Deontological or Utilitarian), the Source (Child or Adult) and Self moral orientation (Deontological or Utilitarian). We predicted that participants would agree more with Utilitarian Adults and agree less with the Deontological Child, according to their own moral orientation. However, there was no interaction between these factors, and therefore the predicted effect was not obtained.

Concerning Participants Moral Judgements in view of our third hypothesis (iii) the predicted pattern of an interaction between Source, Source Answer and Judgement yielded no significant

results. Analysing participants judgments according to their moral orientation, we find that, for Utilitarian participants there was no effect and for Deontological participants the effect was the opposite of what was predicted, as the competent source (Adult) increased the effect of Source Answer, when this Source was Utilitarian.

It is noteworthy that utilitarian participants, give more utilitarian Final Judgements when exposed to the adult utilitarian source, but give more deontological Final Judgements when exposed to the adult deontological source (marginal effect) when compared to their Initial judgements. This suggests that utilitarian participants may be influenced in both directions by the moral orientation of the source perceived as more morally competent.

Another interesting effect that arouse, in spite of not being predicted, was that participants tend to agree more with the Utilitarian Source than the Deontological Source. This effect may be seen as an indication that the participants perceive the Utilitarian Source as more rational or morally sophisticated, triggering more deliberation on participants, leading them to agree more and become more utilitarian in their own judgments. However, this is only a supposition, and further investigation is needed on the subject.

Lastly, most participants in the current study gave an initial utilitarian judgement, in line with Bago & De Neys (2019; see also Mata et al., 2020). Although in the present case we did not asked participants to give the first answer that comes to mind, this result may be seen as further criticism against the strict links between a deontological orientation and intuition, and between a utilitarian orientation and deliberation, suggested by Greene's (2008) model.

In summary, only a few of the predictions were corroborated; some of them, only partially. Participants tend to agree more with the morally competent source than with the incompetent source, and they tend to agree more with the utilitarian source than with the deontological source. When it comes to the influence of the source on participants own moral judgements, utilitarian individuals showed a tendency towards more utilitarian answers after exposition to a utilitarian source; contrary to what was expected, deontological individuals tend to become more utilitarian also. In fact, the tendency to give more utilitarian answers to moral judgements seems to be present for all participants, regardless of the initial moral orientation.

When the Utilitarian participants and Deontological participants are analysed separately, the effects of source competence are evidently present on both groups, but the effect of the source's moral orientation seems to be present only for the Utilitarian participants, as Utilitarian

participants gave more Utilitarian judgements but Deontological participants didn't give more deontological judgements, when exposed to sources that share their own moral orientation. This further reinforces the possibility that Utilitarian sources have a greater impact on individual moral judgement that Deontological sources.

A strong point of this research, I believe, is the use of a baseline (initial judgments) that allowed for the categorization of the participants into Utilitarian and Deontological participants. This baseline allowed for a better evaluation of the impact of others judgments on participants moral judgments. However, this was, simultaneously, a limitation of the present study because the number of initially deontological participants was much smaller than that of utilitarian participants and the number of deontological participants was small in absolute terms. This means that the present findings should be taken as preliminary and that further research (e.g., increasing the sample size) is needed to confirm (or not) these findings.

Another possible limitation this research has is the absence of a measure to understand if the simulated social context worked. In spite of designing the experiment in such a way that the source is believed to be a real person and to include the Bias Blindspot measure, participants may still suspect that the source is not a real person, as there was no control for this issue in the end of the experiment. In other words, there is no way to know if the participants perceived the "others" as more or less morally competent constantly during the study or if it kept being altered in function of the answers of the sources. Finally, we didn't evaluate if the effects found are lasting (e.g., when participants answer again with no source present) or if they are temporary, only occurring when the source is present. In the future, it is intended to provide a set of measures at the end of the study, to control for these matters.

Regarding the answers that the sources gave, only the reactions to the other's extreme answers were explored. According to Mata et al. (2020) these are more intuitive answers, regardless of being extreme deontological or utilitarian. However, there was no way to know if these answers were perceived as such by the participants, which would be an interesting point to explore. In the future, the manipulation should include other tiers of answer extremity (e.g. neutral, moderate, extreme) and measures to assess if the other is perceived as more or less intuitive depending on their answer's extremity.

Additionally, we only tested with one type of dilemma (Trolley Problem dilemma and similar). This may be reductive, and it is necessary to test these effects with other types of dilemmas

(e.g., higher or lower emotional charge), as, for example, dilemmas with a strong personal components (e.g., a relative is present in one of the groups to be sacrificed) has been shown to increase the emotional factor and the rate of deontological responses (Hao et al., 2015). Furthermore, only the impact of judgements was assessed in this study, and no justifications or reasons. According to Haidt (2001), there should be other effects present related to the moral reasoning behind the judgement, and not only to the judgements themselves.

Moreover, only the impact of one source on individual moral judgement was evaluated. It would be interesting and relevant to explore the effects of a group on moral decision-making; specifically, the effects of interpersonal conflict, with situations of several sources with the same moral orientation or opposing moral orientations, minority and majority effects and larger group interactions on individual moral judgement.

Finally, a (*sort of*) limitation is time. Several other study ideas and research lines, including some of the abovementioned limitations, were put on hold due to the time constraints associated with the development of a master's degree research, as there was not enough time to implement them all. Though, as previously stated, these research lines and ideas are merely put on hold, to be implemented in future research.

The achieved outcomes and some of the identified research gaps suggest several new possible research lines which are briefly discussed below:

1) One of the most evident and promising follow-ups is the introduction of a different set of moral dilemmas. Specifically, introducing the idea that one of the people to be sacrificed, either on the larger or on the smaller group, to be a close relative of the participant (e.g. wife, father, mother) in order to increase the rate of deontological answers; on the other hand, the idea of increasing the number of people that can be saved on the larger group, increasing the rate of utilitarian answers. It would be interesting to explore how the social context affects this pattern and what new effects may arise. This study is already being implemented, with research data already collected.

2) A different follow-up is the reversal of the order of the manipulation. Instead of participants seeing the answer of a source, they would answer a moral dilemma and only afterwards see the answer of the source. The next step would be to ask the participant how confident they were of their response, and if they would want to

change their answer, given the answer of the source. This would have the goal of showing that, in spite of their first judgements, individuals can still fall prey to the effects of social conformity and it would be interesting to understand if this changes the moral orientation and if so, to which degree and in what conditions.

3) Finally, an approach consisting in creating an actual social conformity environment, reminiscent of Asch's experiments, where we would put a number of people in a room (e.g., seven other people), where there is only one critical participant and the rest are bogus participants (collaborating with the experimenter). The manipulation would rely on the answers to moral dilemmas in a group environment, where the bogus participants would have instructions to answer in a specific way (Deontological or Utilitarian). Following Asch (1950-56) studies, in a first experiment all participants will consistently give the same response (D or U), in opposition to the critical participant's moral orientation (as obtained in a pre-test). Other manipulations would include, a) the presence of an ally (one of the bogus participants gives the same response of the critical participant); b) varying the number of the majority (e.g., from 3 to 7); c) varying the extremity of the bogus participants answer (e.g., extreme vs. moderate D or U responses). The goal would be to observe the effects that a social group would have on the moral orientation of the participants. The idea is that when in a group, conflict would be resolved more easily if there is a collaborator that shares the participants moral orientation because his presence would reinforce the initial moral judgement, even if there is another collaborator with an opposing view.

5. Final Remarks

There is strong evidence suggesting that social context has an impact in moral judgements and moral decision-making. By showing a significant difference between overall answers when confronted with sources with different moral orientation and moral competence, the results imply that the perceived morality and the perceived competence of the people present in our social environment may change or reinforce our own views and, subsequently, the way we solve moral problems. However, some of the specifics of these interactions are still not clear. This line of research has several different other points that remain to be addressed and the possibilities for further studies are quite promising.

6. References

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*, *70*(9), 1.

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*(2), 381-417.

Bargh, J.A. and Chartrand, T.L. (1999) The unbearable automaticity of being. *Am. Psychol.* 54, 462–479

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral cortex*, *6*(2), 215-225.

Białek, M., & De Neys, W. (2017). Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity. *Judgment and Decision Making*, *12*(2), 148-167.

Butera, F., Legrenzi, P., Mugny, G., & Perez, J. A. (1992). Influence sociale et raisonnement. *Bulletin de psychologie*, *45*(405), 144-154.

Butera, F., Mugny, G., Legrenzi, P., & Pérez, J. A. (1996). Majority and minority influence, task representation and inductive reasoning. *British Journal of Social Psychology*, *35*(1), 123-136.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of personality and social psychology*, *104*(2), 216.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological science*, *17*(12), 1082-1089.

Damásio, A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science*, *264*(5162), 1102-1105.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, *51*(3), 629.

Dolan, R.J. (1999) On the neurology of morals. *Nat. Neurosci.* 2, 927–929.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, *8*(3), 223-241.

Flack, J.C. and de Waal, F.B.M. (2000) 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. In *Evolutionary Origins of Morality* (Katz, L.D., ed.), pp. 1–29, Imprint Academic.

Foot, P. (1967). The problem of abortion and the doctrine of double effect.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25-42.

Freeman, S. (1994). Utilitarianism, deontology, and the priority of right. *Philosophy & Public Affairs*, *23*(4), 313-349.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work?. *Trends in cognitive sciences*, *6*(12), 517-523.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389-400.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144-1154.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.

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

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, *108*(4), 814.

Haidt, J. (2003). The moral emotions. *Handbook of affective sciences*, *11*(2003), 852-870.

Hao, J., Liu, Y., & Li, J. (2015). Latent fairness in adults' relationship-based moral judgments. *Frontiers in psychology*, *6*, 1871.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, *30*(5), 513-541.

Kahane, G., & Shackel, N. (2010). Methodological issues in the neuroscience of moral judgement. *Mind & language*, *25*(5), 561-582.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kant, I., & Quintela, P. (1995). Fundamentação da metafísica dos costumes.

Kohlberg, L. (1969). Stage and Sequence: The Cognitive Developmental Approach to Socialisation, in D. Goslin (ed.), *Handbook of Socialisation Theory and Research*. Chicago: Rand McNally.

Kohlberg, L. (1976). Moral stages and moralization. *Moral development and behavior*, 31-53.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development ', in T. Mischel (Ed), Cognitive Development and Genetic Epistemology, New York, Academic.

Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European journal of social psychology*, *8*(2), 181-192.

Mugny, G., Butera, F., Sanchez Mazas, M., & Pérez, J. A. (1995). Judgements in conflict: The conflict elaboration theory of social influence. *Perception evaluation interpretation. Swiss monographs in psychology*, *3*, 160-168.

Mercier, H., & Sperber, D. (2011). Argumentation: its adaptiveness and efficacy. *Behavioral and Brain Sciences*, *34*(2), 94.

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Mill, J. S., & Bentham, J. (1987). *Utilitarianism and other essays*. Penguin UK

Piaget, J. (1965) *The Moral Judgement of the Child,* Free Press

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*(2), 165-184.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359-393.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal, 94*(6), 1395-1415.

Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, *40*(7), 923-930.

Vega, S., Mata, A., Ferreira, M. B., & Vaz, A. R. (2020). Metacognition in moral decisions: judgment extremity and feeling of rightness in moral intuitions. *Thinking & Reasoning*, 1-18.

7. Supplementary Materials - Section A

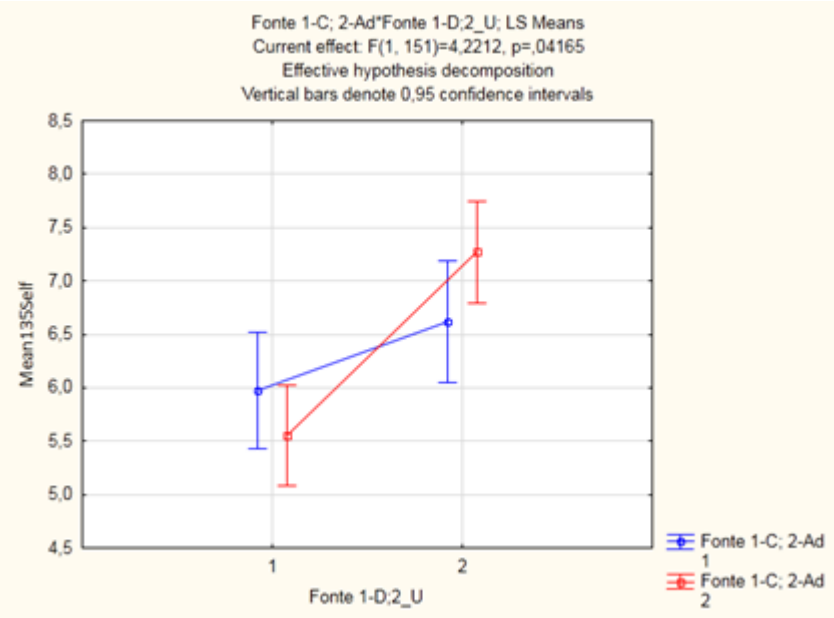

Figure A.1. - 2x2 ANOVA with Source and Source Answer as factor only for the Final Judgements.

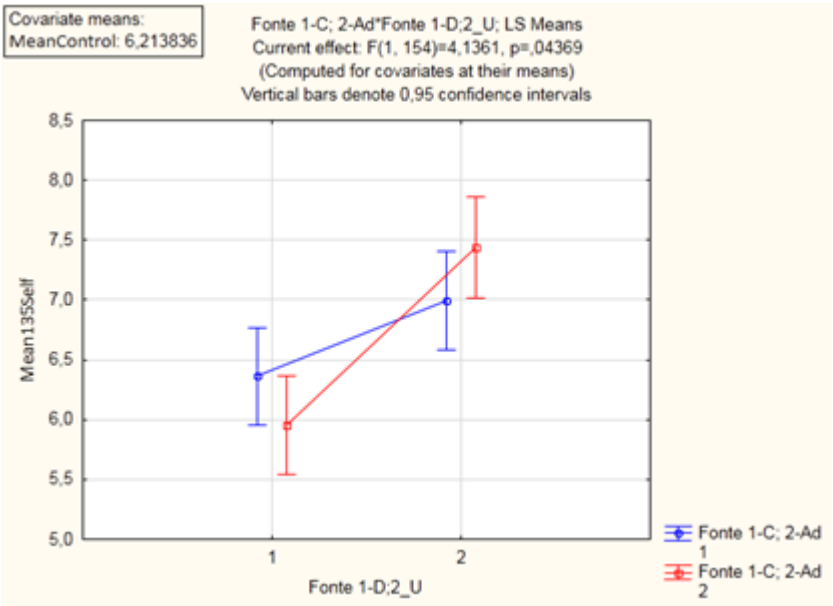

Figure A.2. - 2x2 ANOVA with Source and Source Answer as factor and Initial Judgement as Covariate.

Supplementary Materials - Section B

Dilemmas - Control

“Inside a hazardous materials warehouse are 7 workers doing inventory. Due to a short-circuit, a fire starts that threatens to destroy the building. On the control tower, the foreman sees the fire alarm and has to decide if he should close the fire-doors isolating the fire from the warehouse, directing it to a room where 2 workers are having lunch. If the foreman closes the fire-doors, the 7 workers will be saved, but the 2 inside the room will die. Should the foreman close the doors?”

“A lighthouse keeper notices that a boat carrying 1 crewman is going to colide with a rocky coast. Due to the fog, the crewman can’t see the coast. The lighthouse keeper can warn the crewman to evade the coast and escape safely. However, he notices that there is a second boat that carries 5 people that is also on a colision course with the rocky coast. If the lighthouse keeper warns the first boat, the boat with 5 people will have no chance to evade the coast, will collide and the occupants will die. Should the lighthouse keeper warn the boat with 1 crewman?”

“An airport radar operator notices that there a malfunctioning aircraft is going to make an emergency landing on a strip where 6 workers are working. If the aircraft lands there, the 6 workers will be killed. The radar operator can direct the aircraft to land on a different landing strip instead, where only 2 workers are performing repairs. If the operator instructs the pilot to do this, the 2 workers will be killed, but the other 6 will live. Should the radar operator instruct the pilot to change landing strip?”

Dilemmas - Conflict

“There is a flood near a village with 10 inhabitants. If nothing is done, the whole village will be destroyed, and everyone will die. There is not enough time to move them. Fortunately, an engineer has the power to close a floodgate, diverting the flow of the river towards a lone house, where 2 people live. If he closes the floodgates, the 10 villagers will live, but the 2 people inside the lone house will die. Should the engineer close the floodgates?”

“Due to an accident, 15 miners are trapped on a mining well. They are almost out of oxygen and they will die if nothing is done. However, the supervisor can divert the oxygen from a

nearby well to save the miners. In this well there are 3 miners, that will die if the oxygen is diverted. If the supervisor diverts the oxygen, the 3 miners will die, but the other 15 miners will live. Should the supervisor divert the oxygen?"

"A trolley is barrelling down the railway tracks. Ahead, in the tracks, 5 people are tied and unable to move. The train is going straight at them and there is no way to warn them. However, a railway worker may pull a lever and change the direction of the trolley to another track, where 1 person stands. If the worker pulls the lever, the 1 person will die, but the other 5 will live. Should the worker pull the lever?"

Dilemmas - no-conflict

"Due to a malfunction on a power plant, a leak of a toxic gas threatens to enter a small room where to scientists are working. If nothing is done, they will die from exposure to the toxic gas. The only way to save these 2 workers is for the chief-engineer to press a button that vents the toxic gas to the outside through a chimney. However, he realises that there are 8 construction workers repairing the chimney that will die if the toxic gas is vented through the chimney. Should the engineer vent the gas?"

"There is a radar operation supervising the movement of ships near Greenland. Due toa sudden iceberg displacement, a boat with 3 people is about to hit the iceberg. If nothing is done, all passengers will die. The only way to save them is for the radar operator to order an emergency manouver to change the ship's course. However, this maneuver will hit a life-raft that carries 11 people. If the operator orders the maneuver, the 11 people will die, but the other 3 people will live. Should the radar operator order the maneuver?"

"On a construction site there is a malfunctioning crane that is spinning to the right, uncontrollably. Close to the crane there is a building with 6 workers and the weight of the crane is going straight for them. If nothing is done, the crane will destroy the building and kill the 6 workers. However, the crane operator can pull a lever and change the direction of the rotation to the left, where there is only and empty building. If the crane operator pulls the lever, the crane will destroy the empty building, killing no one and saving the 6 workers. Should the crane operator pull the level?"

Supplementary Materials - Section C

Source's Morality Problem (Thief Problem)

"John is going down a street on a deserted Lisbon suburbia when he sees, a few meters ahead, an elderly woman who drops her wallet and goes insider her house. Inside the wallet is the elderly woman's retirement money that she just received, totalling 800€ in bills. John picks up the wallet and counts the money. Nobody sees him, there is no surveillance whatsoever and no kind of technology on the wallet allows its location to be known. John decides to keep the wallet and the money, and to leave that place as fast as possible."