Theory and Society (2024) 53:1441–1472
https://doi.org/10.1007/s11186-024-09584-1

Class-based differences in moral judgment: A bayesian approach

Andreas Tutić[1]

Accepted: 25 October 2024 / Published online: 25 November 2024

Abstract

This study employs Bayesian inference to explore class-based differences in moral judgment. Based on the dual-process perspective in interdisciplinary action theory, we estimate in a first step a process model which differentiates parametrically between emotionally driven deontological, deliberatively driven utilitarian, and residual judgmental inclinations. In a second step, our estimates of these parameters are correlated via beta regressions with indicators of social class and thinking dispositions. We find a considerable association between social class, specifically income, and deontological inclinations, whereas consequentialist inclinations correlate with thinking dispositions but not with social class. This research underscores the utility of Bayesian estimation in closing the gap between theoretical and statistical modelling. Employing this approach enhances our understanding of the nuanced interplay between intuitive and deliberative processes in moral judgment and, more generally, offers a promising direction for advancing sociological action theory.

Keywords Dual process · Moral dilemmas · Bayesian estimation · Mediation · Thinking dispositions

Introduction

In the evolving landscape of interdisciplinary action theory, the investigation into the foundations and expressions of moral judgment continues to be of paramount interest. At the heart of this inquiry lies the distinction between utilitarian and deontological ethics, principles that guide individuals towards outcome-focused or duty-bound moral reasoning, respectively (Kant, 1797; Mill, 1863; Foot, 1967; Thomson, 1986). These theoretical underpinnings serve not only as a bedrock for philosophical discourse but also as a lens through which the complexities of human moral decision-making are examined (Greene et al., 2001; Conway & Gawronski, 2013; Tutić

✉ Andreas Tutić
andreas.tutic@uib.no

[1] Sosiologisk institutt, Universitetet i Bergen, Serviceboks 7802, 5020 Bergen, Norge

et al., 2024). The dynamic interplay between utilitarian and deontological inclinations shapes our understanding of moral judgments across various social and individual contexts, highlighting the need for nuanced approaches in capturing the essence of moral reasoning (Haidt, 2001, 2003; Greene et al., 2001, 2004, 2008; Tutić et al., 2022).

Building upon the foundational work of scholars such as (Greene et al., 2001, 2004, 2008) and Conway and Gawronski (2013), this paper seeks to extend the exploration of moral judgment through the application of Bayesian statistics. Previous studies have employed the psychometric technique of process dissociation to disentangle utilitarian and deontological judgment tendencies (Jacoby, 1991; Payne & Bishara, 2009). Broadly speaking, process dissociation uses parameterized models that link observable behavior to underlying propensities to engage in intuitive Type 1 or reflective Type 2 processing. These models are constructed such that when individuals are confronted with a suitably designed series of judgment problems, researchers can infer, based on behavior, the underlying tendencies to engage in either Type 1 or Type 2 processing. Process dissociation is particularly valuable in the study of moral judgment, where it helps quantify the degree to which utilitarian and deontological inclinations contribute to moral judgment (Conway & Gawronski, 2013; Conway et al., 2018; Tutić et al., 2024). Unlike previous studies that employed process dissociation, our study uses the analytical power of Bayesian methods (Kruschke, 2015; McElreath, 2020) to study moral judgment. This shift is motivated by the growing recognition of Bayesian statistics' utility in providing a more flexible and nuanced framework for the estimation of action-theoretical process models. The reliance on prepacked statistical routines such as variants of the generalized linear model often necessitates to reduce well-developed theoretical models to mere qualitative implications regarding associations of key variables. The substantive gap between theoretical and statistical models not only hinders scientific progress because the relationship between qualitative hypotheses and theoretical process models is lose at best, since the same hypothesis can usually be derived from a multitude of theoretical models (e.g. McElreath, 2020: 6). The mismatch between theoretical and statistical models also disincentivizes investments in proper theoretical modelling since the intricacies of these models often eschew empirical testing (Leamer, 1983; Gelman & Shalizi, 2013). Bayesian methods provide the flexibility to estimate nuanced models that can accommodate the multifaceted nature of social action and judgment, transcending the limitations imposed by traditional statistical approaches.

In this vein, the current paper endeavors to contribute to the rich tapestry of the study of moral judgment by reexamining its relationship with social class through a Bayesian lens. Drawing inspiration from the seminal work of (Côté et al., 2013), which presents evidence supporting the hypothesis that individuals from higher social classes exhibit a stronger tendency towards utilitarianism compared to their lower-class counterparts, we seek to delve deeper into this phenomenon. Building upon these findings, our study revisits the data analyzed by Tutić et al. (2024), which employed process dissociation to explore the same relationship. While Tutić et al. (2024) report a relatively weak direct association between utilitarianism and social class, they do find that higher-class actors tend to exhibit lower levels of

deontological judgment than their lower-class counterparts. This distinction underlines a nuanced aspect of the relationship between social class and moral judgment, emphasizing the reduced adherence to deontological principles among individuals from higher social strata. In response to these insights and the methodological and theoretical limitations associated with process dissociation, our paper adopts a Bayesian statistical approach to estimate an alternative dual-process model of moral judgment. As will be shown in more detail in the theory section, this model not only addresses key methodological shortcomings of the process dissociation approach (Batchelder & Riefer, 1999; Payne & Bishara, 2009; Miles et al., 2023), such as hard-to-interpret estimates for the U parameter and the neglect of inherent uncertainty in parameter estimates. The alternative model also corrects an oversimplification in the process dissociation model, which relates to the absence of an error term in the model specification, and aligns more closely with the default-interventionist framework of general dual-process theory (Evans, 2010; Kahneman, 2011; Stanovich, 2011) and the dual-process perspective in cultural sociology (Kroneberg, 2005, 2011; Esser & Kroneberg, 2015; Tutić, 2022; Vaisey, 2009; Miles, 2015; Lizardo, 2017). Specifically, in line with default-interventionism, the alternative model specifies deontological snap judgments as the default, which may be corrected by more reflective utilitarian reasoning. Through this approach, this paper contributes both in theoretical as well as methodological terms to our understanding of class-based differences in moral judgment.

The remainder of this paper is structured as follows: Sect. 2 provides the action-theoretical basis for the dual-process of moral judgment which is used in this study. In Sect. 3 we describe the sample, the variables, and provide the details regarding our Bayesian estimation strategy. Section 4 presents the empirical findings of this investigation. Finally, Sect. 5 concludes with a summary of our findings, a discussion of their relationship to the literature, a discussion of the limitations of this study, as well as suggestions for future research.

Theory and previous research

Dual-process models of moral judgment

The field of behavioral science frequently employs moral dilemmas—hypothetical situations that juxtapose deontological against utilitarian ethics—to explore the mechanics of moral judgment (Petrinovich et al., 1993; Greene et al., 2001, 2004, 2008; Nichols, 2002; Mendez et al., 2005). These scenarios are instrumental in revealing the nuanced interplay between the principles that guide ethical decision-making, highlighting the conflict between adherence to deontological concepts of moral rights and duties and the outcome-based reasoning characteristic of utilitarian thought.

In making sense of the empirical findings on moral dilemmas and therefore in understanding the complex interplay between deontological and utilitarian inclinations in moral judgment, recent discourse centers on Greene's dual-process theory of moral reasoning, which distinguishes between emotion-driven deontological

judgments and the more deliberative utilitarian judgments (Greene et al., 2001; Gree ne, 2007, 2013).[1] This theory posits that emotional impulses typically underlie deontological thinking, whereas utilitarian decisions are informed by more analytical thought processes. Empirical investigations, including studies on individuals with brain injuries (Mendez et al., 2005; Ciaramelli et al., 2007; Koenigs et al., 2007), have lent consistent support to the idea that deontological judgments are predominantly fueled by emotional responses. Research has further illuminated how strategies for regulating emotions (Lee & Gino, 2015; Li et al., 2017) and challenges associated with managing emotions (Zhang et al., 2017) distinctly affect deontological tendencies without exerting similar effects on utilitarian thinking.

In the study of moral judgment, the concept of process dissociation (Jacoby, 1991; Kelley & Jacoby, 2000; Yonelinas, 2002; Payne & Bishara, 2009) offers a nuanced framework for understanding how individuals navigate complex ethical decisions (Conway & Gawronski, 2013; Conway et al., 2018; Fleischmann et al., 2019). This approach involves presenting subjects with moral dilemmas designed to tease apart their leanings towards either utilitarian or deontological decision-making principles. Central to this methodology is the categorization of dilemmas into congruent and incongruent types, based on the anticipated moral evaluations from adherents of both ethical perspectives. In a congruent dilemma, actions deemed morally wrong by both utilitarian and deontological standards are presented, suggesting a universal ethical stance against the proposed behavior. Conversely, in incongruent dilemmas, a divergence emerges: utilitarian-oriented individuals may find certain actions morally acceptable due to the greater good they serve, while those with a deontological orientation may condemn these behaviors due to the inherent harm they cause.[2]

The process dissociation model of moral judgment, proposed by Conway and Gawronski (2013), aims at quantifying the inclination towards utilitarian (U) and deontological (D) judgment based on observed judgment in a series of congruent and incongruent dilemmas (see Fig. 1). More specifically, let p_con be the proportion of congruent dilemmas in which a respondent judged the described action as immoral and let p_incon be the respective proportion of incongruent dilemmas. Then, according to the process dissociation model of moral judgment, $p\_con = U + (1 - U)D$ and $p\_incon = (1 - U)D$ and, therefore, $U = p\_con - p\_incon$ and

[1] While we believe that dual-process theories provide a valuable and fruitful perspective in interdisciplinary action theory, and in particular in the study of moral judgment, the approach has also been subjected to serious criticism (e.g., Osman, 2004; Kruglanski & Gigerenzer, 2011; Melnikoff & Bargh, 2018). One major point of criticism refers to the fact that the dichotomies used to characterize Type 1 and Type 2 processes (such as fast vs. slow or autonomous vs. controlled) are empirically not as aligned as suggested in the dual-process literature. Another important point of criticism argues that many of the empirical phenomena commonly cited as evidence for a Type 1 / Type 2 dichotomy can also be explained by more parsimonious single-process accounts. Needless to say, these points of criticism have been contested by proponents of the dual-process approach (e.g., Evans & Stanovich, 2013). While we do not directly address the ongoing theoretical controversy in this paper, we believe that cumulative evidence from applied studies such as this will prove useful in clarifying the contested issues.

[2] See the supplementary materials for a presentation of the set of congruent and incongruent dilemmas used in this study.

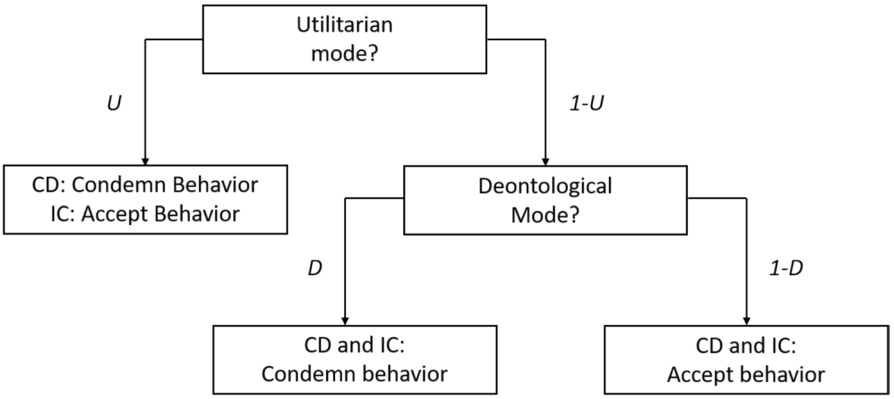

Fig. 1 The process dissociation model of moral judgment (CD: congruent dilemma, ID: incongruent dilemma)

D = p_incon / (1 – U). Note that U takes values between −1 and 1 and D varies between 0 and 1.

Empirical studies making use of the process dissociation model have consistently demonstrated that the differentiation between two types of moral inclinations leads to both sociologically and psychologically illuminating as well as theoretically sensible results. First and foremost, contrary to conceptions according to which are greater inclination towards utilitarianism goes necessary hand in hand with a smaller inclination towards deontology (e.g. Côté et al., 2013), empirical research using process dissociation typically finds only weak negative and also insignificant correlations between U and D (Conway & Gawronski, 2013; Conway et al., 2018; Fleischmann et al., 2019; Tutić et al., 2024). Further, empirical research has uncovered that U and D are differently attached to various covariates such as concerns about following moral rules or certain reasoning processes (Piazza & Landy, 2013; Reynolds & Conway, 2018). A sociologically interesting finding in this regard is that social class is more strongly associated with D than with U (Tutić et al., 2024). Hence, the process dissociation model represents a theoretically sophisticated and empirically fruitful further development of the dual-process perspective in the field of moral judgment.

The process dissociation model offers valuable insights into moral judgment, but it also suffers from several key limitations. First, it assumes a hierarchy between utilitarianism and deontology, where utilitarian judgment is considered first, and only if it fails, deontological reasoning is engaged. This structure contrasts with dual-process theories (Greene et al., 2001; Kahneman, 2011) and default interventionism, which suggest that deontological reasoning is an intuitive, default Type 1 process, whereas utilitarian reasoning requires reflective, Type 2 processing. Second, the model struggles with negative U values that arise when p_con is smaller than p_incon, leading to problematic interpretations of utilitarian tendencies. These cases violate the assumption that utilitarian judgments represent a non-negative probability, complicating the interpretation of the data. Third, the process dissociation

model assumes that if neither utilitarian nor deontological processes are engaged, the behavior will not be condemned. This assumption is an oversimplification since real-world moral judgments often involve additional motivations beyond purely deontological or utilitarian inclinations. Some individuals may still morally condemn behavior for reasons not captured by either of these two constructs, such as personal biases or contextual factors. Finally, from a statistical perspective, the algebraic estimation strategy used in the process dissociation model does not account for the inherent randomness in the decision-making process or offer information about the uncertainty of estimates. The model provides point estimates for U and D, which can give the false impression of precision. A more robust approach would take into account both parametric and sampling uncertainty.

Against this background, in this paper we make use of an alternative dual-process model of moral judgment which is depicted in Fig. 2 (see also the related literature on multinomial processing tree models; Batchelder & Riefer, 1999; Payne & Bishara, 2009; Miles et al., 2023). We will briefly describe this alternative model before explaining how it overcomes the deficiencies of the process dissociation model outlined above. According to this model, a decision maker faced with a moral dilemma is characterized by three parameters, as, rc, and epsilon, which all are (conditional) probabilities and take values in the unit interval. The as-parameter, in which as stands for automatic-spontaneous mode (Kroneberg, 2005), gives the probability that the decision maker comes to a deontological judgment based on a hot and fast Type 1 process. This means that the decision maker finds the behavior described in both congruent and incongruent dilemmas too harmful to be morally acceptable. Given that this is not the case, with probability rc, which stands for reflecting-calculating mode (Kroneberg, 2005), the decision maker responds with an utilitarian judgment based on a cold and slow Type 2 process. Accordingly, the decision maker will condemn the described behavior as too harmful in the congruent dilemmas, but find the behavior in the incongruent dilemmas to be morally

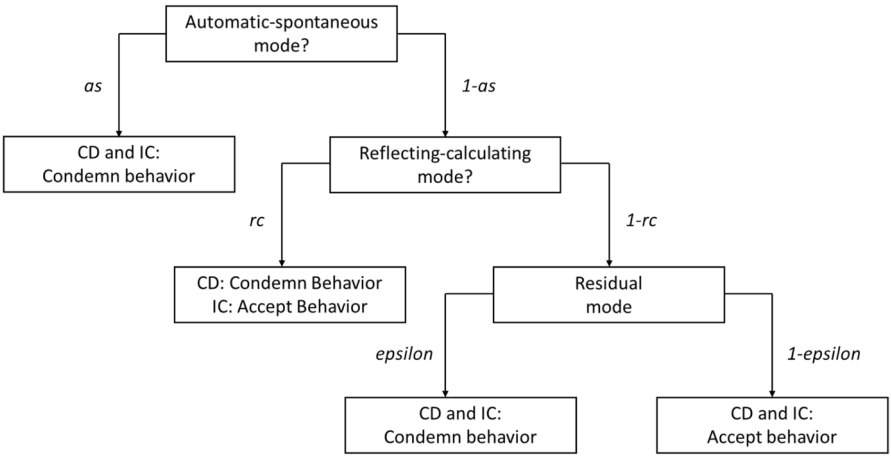

Fig. 2 The DP-model of moral judgment (CD: congruent dilemma, ID: incongruent dilemma)

acceptable. Finally, if the moral judgment is neither driven by the as- or the rc-mode (an event with probability (1 – as)(1 – rc)), it is assumed that the decision maker with probability epsilon will condemn the behavior described in the moral dilemma, and with probability 1 – epsilon accept the behavior. Epsilon can be thought of as an individual disposition to be morally judgmental in the absence of both decisive deontological intuitions or decisive utilitarian reasons.

A key advantage of the DP-model over the process dissociation model is its ability to resolve several limitations inherent in the latter. First, while the process dissociation model suffers from algebraic issues, particularly in situations where the utilitarian parameter (U) can take negative values, the DP-model introduces three parameters—automatic-spontaneous mode (as), reflecting-calculating mode (rc), and epsilon—which avoid this issue entirely. These parameters are structured as probabilities within the unit interval, ensuring that no negative values arise, and offering a more coherent representation of moral judgment. Second, the DP-model employs Bayesian estimation methods, addressing the lack of uncertainty quantification in the process dissociation model. Whereas the process dissociation model provides point estimates without insight into the imprecision of those estimates, Bayesian methods in the DP-model allow for estimates that take into account both parametric and sampling uncertainty, providing a more robust and nuanced analysis of moral inclinations. Third, the DP-model introduces the epsilon parameter, which captures individual variability in moral judgments beyond utilitarian or deontological reasoning. This addition resolves the oversimplification present in the process dissociation model, where judgments were assumed to only stem from either utilitarian or deontological considerations. Finally, the DP-model aligns more closely with the broader dual-process perspective by giving priority to the fast, intuitive Type 1 processes (deontological reasoning) over reflective Type 2 processes (utilitarian reasoning).

Social class and moral judgment

The idea that moral judgment differs between social classes is suggested by various theoretical approaches that cut across disciplinary boundaries in sociology and social psychology. First, research on contextualism and solipsism is essential for understanding how social class might influence moral reasoning and the conception of self. These concepts help explain how different levels of material resources shape the ways individuals relate to others and to their social environments. Contextualism refers to an orientation in which individuals from lower social classes focus on external forces and situational constraints, given their reliance on social networks and environmental factors for survival (Kraus et al., 2012). In contrast, solipsism characterizes individuals from higher social classes, who are more attuned to their internal states, personal goals, and autonomy due to the greater control they have over their lives (Kraus & Keltner, 2009). This distinction reflects fundamental differences in how social classes navigate and interpret their worlds, rooted in material resources and social positioning. Lower social classes—facing more constrained environments—are generally more focused on interdependence and external social

cues. As a result, they develop a communal orientation in which the needs of others and social cohesion are central to their identity. This orientation leads to behaviors that emphasize empathy, collective welfare, and shared responsibilities, which align with contextualism. In contrast, higher social classes—having more access to resources—are oriented toward solipsism. They view themselves as autonomous agents, free from many of the external constraints that shape the lives of lower-class individuals. This agentic conception of the self emphasizes independence, self-reliance, and personal achievement. In terms of moral judgment, two specific aspects of the theorized association between class and contextualism/solipsism are particularly instructive.

One significant aspect of this class-based difference in orientation is reflected in how individuals interpret social relationships. Lower-class individuals, with fewer resources and a greater need for support from others, are more likely to adopt a communal understanding of relationships, where behavior is driven by the needs of the involved actors and generalized reciprocity, rather than by expectations of direct reciprocity (Ekeh, 1974; Clark & Mills, 1993; Uehara, 1990). Communal relationships align with the norms of indirect reciprocity, where help is given with the understanding that it will eventually be reciprocated, either by the recipient or another community member. On the other hand, higher-class individuals, who have more control over their resources and environments and are less dependent on others, tend to interpret relationships as exchange relationships. Exchange relationships are more transactional, reflecting the agentic self-conception of higher-class individuals, who prioritize autonomy and independence over communal obligations (Piff et al., 2012). These exchange relationships are characterized by direct reciprocity, where the giving of a good is contingent upon receiving something of equal value in return (Clark & Mills, 1993). In the domain of social exchange, higher-class individuals tend toward a utilitarian, calculating stance, characterized by mental accounting and a "quid pro quo mentality" (Uehara, 1990: 526). Direct empirical evidence for these claims is provided by Tutić and Liebe (2019), who show through different versions of the dictator game that higher-class actors have a greater affinity for direct reciprocity, whereas lower-class actors are more oriented toward indirect reciprocity. Given this background and the ideas regarding the chronic availability of frames and goal frames as mental models in newer sociological action theories (Esser & Kroneberg, 2015; Lindenberg, 2008), it can be argued that higher-class actors are expected to adopt a similar stance in moral judgment and therefore tend toward utilitarian judgment.

A second relevant aspect of class-based differences in solipsism and contextualism relates to the concept of empathy (Batson, 2011), the capacity to feel and understand the emotions of others. Individuals from lower social classes, who are more likely to experience precarious and unpredictable environments, tend to develop higher levels of empathy as an adaptive strategy for maintaining social bonds and managing external pressures (Kraus et al., 2012; Kraus & Keltner, 2009). According to this argument, empathy is strongly influenced by the income-related aspects of social class. Lower-income individuals, due to their reliance on social networks for survival, tend to cultivate stronger empathic tendencies compared to wealthier individuals, who experience fewer existential risks and have more control over their

environments (Côté et al., 2013; Kraus et al., 2012). Research by Côté et al. (2013) suggests that social class differences in moral reasoning may be partly attributed to variations in empathic concern. Empathy is particularly relevant for deontological reasoning, which is often driven by emotional responses to moral dilemmas. Greene's dual-process model (2001) posits that deontological judgments are primarily the result of Type 1 processes—fast, intuitive, and emotionally charged. Higher levels of empathy make individuals more sensitive to the immediate emotional impact of moral dilemmas, encouraging them to prioritize avoiding harm (Batson, 2011). This reasoning suggests that lower-class actors, being more empathic, have a greater tendency toward deontological judgment than higher-class actors.

In addition to the roles of contextualism and solipsism in shaping moral judgment across social classes, we can further explore these differences through the lens of interaction ritual theory, as proposed by Collins (2004). Drawing on Durkheim's work on religious solidarity (Durkheim, [1915] 2008) and Goffman's interactionism (Goffman, 1967), Collins develops a theory of group solidarity that explains how actors build moral solidarity through participation in interaction rituals. These rituals involve physical co-presence, shared emotional experiences, and a common focus of attention, all of which contribute to the formation of collective beliefs and values. The intensity of the emotional energy generated by such rituals fosters moral solidarity, which aligns participants around common cultural norms and beliefs. Collins (1988) extends this theory to argue that social class influences the types of interaction rituals individuals typically engage in, and thus the forms of solidarity they experience and the type of culture they embrace. Specifically, higher-class actors tend to participate in more socially diverse interaction rituals, meaning they interact with a wider variety of people and encounter a broader range of cultural ideas and beliefs (Pichler & Wallace, 2009; Tutić & Liebe, 2020). This social diversity fosters open-mindedness and a more universalistic outlook, as well as generalized trust (Hooghe et al., 2009), as actors are exposed to differing perspectives and learn to relativize their own cultural beliefs. In contrast, lower-class actors typically engage in more homogeneous interaction rituals, restricted to long-standing relationships within tight-knit communities, such as family, neighbors, or close friends. Lacking in diversity of interaction rituals, lower-class actors are prone to develop particularistic cultures, which are treated as reified and enacted in a taken-for-granted fashion. Cultural differences between lower-class and higher-class actors are, of course, a classical theme in sociology (Bourdieu, 1987). For instance, "in the realm of aesthetic tastes, the higher social classes hold reflexive and relativistic ideas as to what constitutes 'art' while the lower classes reject intellectualized art in favor of what seems unreflectingly pretty, sentimental, or colorful" (Collins, 1988: 217). While Collins does not explicitly address the issue of moral judgment, we argue that this line of reasoning naturally leads to the idea that classes differ in moral judgment because of class-based differences in the diversity of interaction rituals. That is, we argue that the rather unreflected, taken-for-granted nature of lower-class culture goes hand in hand with deontological judgments based on Type 1 intuitions. In contrast, utilitarian judgment, which stems from the reflective application of an abstract and universal moral principle (the greatest good for the greatest number), is more akin to the intellectualized cultures of higher classes.

While the concepts of contextualism and solipsism and the framework of interaction rituals are certainly illuminating regarding class-based differences in moral judgment, the current paper—which makes use of the dual-process perspective to study the relationship between social class and moral judgment—focuses primarily on cognitive styles as the central mediating mechanism (Tutić et al., 2024). Cognitive styles refer to the habitual ways in which individuals process information and make decisions. The dominant approach within the dual-process perspective conceptualizes the interplay of Type 1 and Type 2 processes along the lines of default interventionism (Evans & Stanovich, (Evans & Stanovich, 2013). Accordingly, Type 1 processes automatically provide one (or more) default responses in a given choice situation, while Type 2 processes kick in to potentially override these defaults. The level of Type 2 engagement depends on both motivational factors and cognitive resources (Evans, 2018). Cognitive styles, as measured by the Cognitive Reflection Test (Frederick, 2005) and the Faith in Intuition Scale (Epstein et al., 1996), figure as important dispositional motivating factors for engaging in reflective Type 2 processing. Following Greene's dual-process model (2001), we expect deontological judgments to result from "hot" emotional reactions provided by Type 1 processes, whereas utilitarian judgments stem from the "cold" reflective processes of Type 2. Accordingly, we expect utilitarian judgment to correlate with reflective thinking dispositions, whereas deontological judgment is more prevalent among actors who rely on intuitive processes. Importantly, social classes can be expected to differ in cognitive styles. Higher social classes, with their greater access to education and cultural capital, are more likely to develop Type 2 cognitive styles, which favor reflective reasoning (Brett & Miles, 2021). Educational systems catering to the upper class emphasize critical thinking, problem-solving, and abstract reasoning, all of which align with the deliberate, reflective decision-making associated with Type 2 processes (Bourdieu, 1987; Dewey, 1933). In contrast, individuals from lower social classes, who often face economic hardship and uncertainty, are more likely to rely on Type 1 processes, which are intuitive and automatic (Mani et al., 2013). Cognitive styles thus provide a robust mechanism for explaining class-based differences in moral reasoning. As Tutić et al. (2024) demonstrate, higher-class individuals, with their greater reliance on reflective cognitive styles, are more likely to engage in utilitarian judgments, while lower-class individuals, whose cognitive styles are more intuitive and emotionally driven, tend to favor deontological judgments.

Hypotheses

In this study, we will reanalyze the data by Tutić et al. (2024) but instead of estimating U and D based on the process dissociation model, Bayesian statistics will be employed to estimate the DP-model of moral judgment depicted in Fig. 2. Our exposition will focus on two kind of questions: Firstly, we assess whether Bayesian estimation of the parameters of the DP-model yields results that are theoretically coherent. Secondly, we explore the implications of utilizing the DP-model in understanding variations in moral judgment across different social classes.

We can address the first question by testing whether our estimates of as and rc covary with well-established indicators of thinking dispositions such as the Cognitive Reflection Test (Frederick, 2005) and the Faith in Intuition Scale (Epstein et al., 1996):

H1. *Indicators of more intuitive thinking dispositions should be positively associated with as and negatively associated with rc. Indicators of more reflective thinking dispositions should be negatively associated with as and positively associated with rc.*

Regarding the second question, we will check for covariation between indicators of social class such as education and income and the DP-parameters. In an attempt to replicate the findings of both Côté et al. (2013) and Tutić et al. (2024), we can formulate the following hypotheses:

H2. *Indicators of higher social class should be negatively associated with as.*
H3. *Indicators of higher social class should be positively associated with rc.*

Finally, against the background that there are class-based differences in thinking dispositions (Brett & Miles, 2021) and the findings of Tutić et al. (2024), we are also interested in the question of whether observed differences in moral judgment between classes are mediated by class-based differences in thinking dispositions. We can formulate this interest as a weak hypothesis as follows:

H4. *The association between indicators of social class and as and the association between indicators of social class and rc depend on whether or not we control for indicators of thinking dispositions.*

Note that we do not formulate hypotheses regarding the third DP-parameter. While we do believe that this moral construct is interesting in itself as a measure of being morally judgmental, in this study it serves as a kind of residual error term, which we primarily estimate to purify the estimates of as and rc.

Methods

Sample and variables

The study was conducted using an online survey using respondi, a German platform that recruits volunteers for various opinion polls. The participant pool was drawn from a randomly selected group of 13,591 individuals from the panel, with 3,465 initially responding to the outreach. Of these, 2,646 participants completed the survey in full, agreed to participate, and passed a critical quality assurance question, making their responses eligible for our analysis. After employing listwise deletion for missing data we arrived at 2,455 participants. 184 of the missing cases stem from missing responses or responses coded as missing to the open question regarding the participants' income and 22 missing cases stem from missing responses to one of

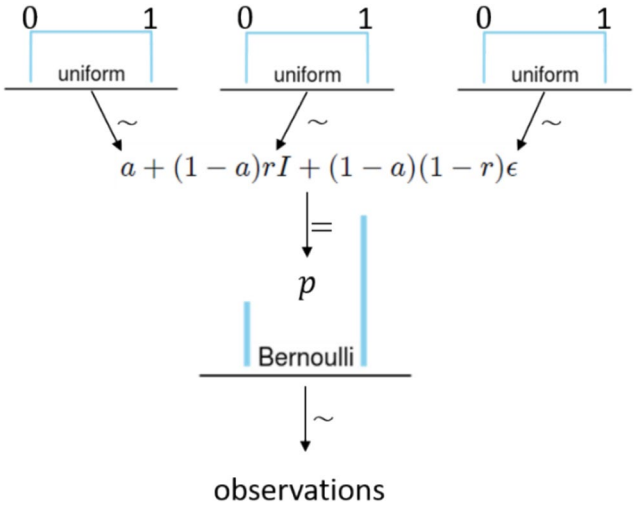

Fig. 3 Kruschke diagram regarding the estimation of the DP-parameters (first stage)

the moral dilemmas; some participants had missings in both income and some moral dilemma.[3]

Adopting the approach by Conway and Gawronski (2013), participants were presented with five morally charged scenarios (covering topics such as abortion, car accidents, vaccine policies, animal research, and border crossing; see the supplementary material) in both congruent and incongruent dilemma versions, resulting in a total of ten dilemmas. These dilemmas were translated into German and displayed to participants in a randomized sequence across different screens. From the observed judgment in these ten dilemmas, we derived the estimates for the DP-parameters as, rc, and epsilon on the individual level.

In addition, this study uses two indicators of social class, two indicators of thinking dispositions, and a couple of sociodemographic controls.

Regarding social class, we rely on variables related to income and educational attainment. The variable "Income" is based on an open question regarding the participants' disposable household income. In an attempt to correct for survey trolling, we set unrealistically high (> 10.000 €) and unrealistically low (< 432 €) answers to missing values; these cutoff points were chosen after graphically inspecting the distribution of the variable and looking for discontinuities in the tails. After this correction, the variable has a mean of 2792.81 and a standard deviation of 1546.93. For the assessment of educational attainment among participants, a categorization scheme adapted from the European Social Survey was employed, which is well-suited to

[3] While Bayesian estimation allows to include powerful methods of imputation (McElreath, 2020: 499–516), we opted against these extended models. While our results were qualitatively identical, the models including imputation had imputed values for p_con and p_incon that were not multiples of 0.2, thereby making the neat graphical display of our results (e.g. Fig. 3) impossible.

capture the nuances of Germany's dual education system. This system integrates academic education with vocational training. The education variable categorizes participants into three tiers, based on their level of educational achievement. The first tier ("Low education") includes individuals without any specialized vocational training or academic qualifications that would enable entrance into higher education (38.86%). The second tier ("Middle education") comprises those who have completed some form of specialized vocational training or possess qualifications for limited access to higher education (35.44%). The third and highest tier ("High education") is reserved for individuals who have undergone extensive vocational training or achieved a higher education degree (26.12%).

With respect to thinking dispositions, we rely on two established instruments from the dual-process literature (see the supplementary materials for details). The Cognitive Reflection Test consists of three questions that are meant to induce subjects into relying on intuitive answers which are in fact wrong (Frederick, 2005). We use a binary indicator "CRT" that takes value 1 if the subjects answered correctly to at least 2 of these questions (35.11%). The variable "FI" is the Faith in Intuition Scale (Epstein et al., 1996), i.e., it is an additive index over 15 items using a scale from 1 to 7 ($\alpha = 0.86$). FI has a mean of 4.73 and a standard deviation of 0.90.

Turning to controls, the variable "Age" quantifies the age of participants in years, with an average age of 49.74 and a standard deviation of 15.72. The "Female" variable is binary, coded as 1 for female participants, who make up approximately 50.6% of the sample. The "GDR" variable is another binary indicator, reflecting whether a participant resides in the eastern part of Germany, with about 21.2% of respondents falling into this category. The "Urban" variable is another dummy, coded as 1 for individuals living in or near large urban centers which is true for 40.2% of the study population. Lastly, the "Couple" variable, also binary, identifies respondents who are married and cohabitating with their spouse, encompassing nearly half of the participants at 50.5%.

Income, FI, and Age were z-standardized before included in the analysis.

Bayesian estimation

As indicated, in this study we make use of Bayesian statistics. In general, Bayesian methods offer several distinct advantages over traditional frequentist statistics. One of the main benefits is the flexibility to model data more accurately, without relying on restrictive assumptions such as homogeneity of variances or normally distributed errors. Additionally, Bayesian analysis provides richer inferences by generating full posterior distributions, offering a more nuanced understanding of uncertainty. Unlike frequentist methods, Bayesian inference does not depend on p-values or sampling distributions, enabling more intuitive and direct interpretations of parameter estimates. This flexibility and clarity in handling uncertainty make Bayesian methods particularly well-suited for complex models (Kruschke, 2015). Both of these general advantages directly apply to the study at hand. First, it is the flexibility of the Bayesian framework that allows us to estimate the DP model of moral judgment without making compromises for statistical identification. This methodological

choice underscores the potential for sociological action theory to gain from combining parametric decision models with Bayesian estimation, effectively bridging the gap between theoretical reasoning and statistical modeling. Additionally, the interpretation of our findings is greatly enhanced by the availability of posterior distributions for the DP parameters, allowing for a more detailed and accurate representation of uncertainty in our estimates.

Our statistical analysis will proceed on two stages. On the first stage, we estimate the DP-parameters. On the second stage, we estimate a beta regression for each DP-parameter to identify covariates of these parameters. Figures 3 and 4 present Kruschke diagrams for both stages (Kruschke, 2015).

The model on the first stage is estimated for each individual separately. Recall that each individual makes 10 binary judgments in 10 moral dilemmas. We model these observations as being sampled from a Bernoulli distribution with underlying parameter p. Following the DP-model of moral judgment, this parameter p in turn is modelled as a function of underlying DP-parameters as, rc, and epsilon as well as an indicator I that takes 1 if the dilemma at hand is congruent. Note that this statistical model is not merely similar but in fact identical to the decision-theoretical model we motivated and explained in the theory section. Finally, regarding priors, we assume that all three DP-parameters are uniformly distributed on the unit interval. This seems to be the only sensible choice regarding the priors, since we had no clue

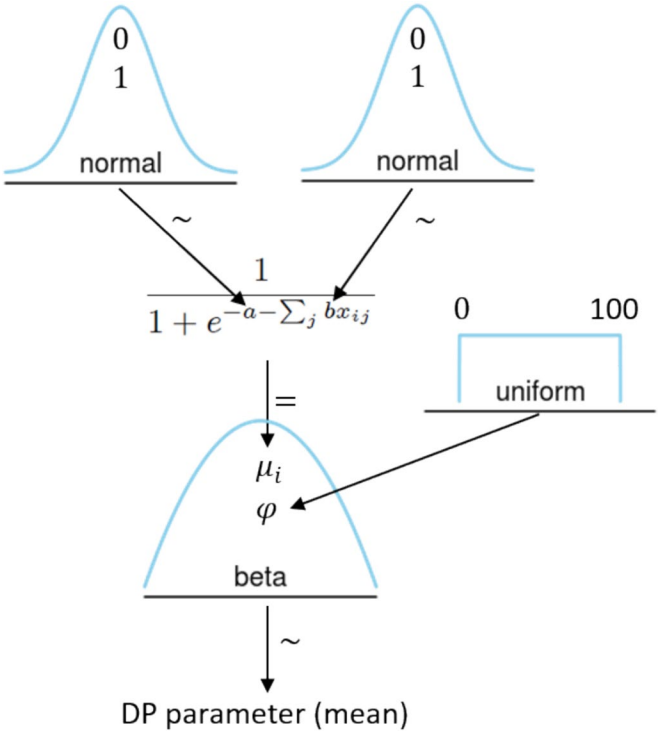

Fig. 4 Kruschke diagram regarding beta regressions (second stage)

what to expect regarding these estimates prior to conducting this research. The first stage results in posterior distributions for each DP-parameter and each individual. We use the mean of the individual-level posterior as our estimate for the respective individual-level DP-parameter.

On the second stage, we estimate a beta regression for each of the DP-parameters using the indicators for social class and thinking dispositions as well as the controls as predictors; these regressions are run over all individuals in the sample. Beta regressions are appropriate because these dependent variables take values in the unit interval. We model the estimates of the DP-parameter as resulting from draws out of the beta distribution. The beta distribution can be parameterized by a central tendency mu and a dispersion parameter phi. The central tendency mu is modelled as a function of the individual level covariates, i.e., indicators of class, thinking dispositions, and controls, where we follow the convention to use the logistic function to map from the reals into the unit interval. Put differently, we model the logodds of mu as a linear function of the predictors. Note that on stage 2 we make use of the generalized linear model framework, not because of limitations in the statistical possibilities, but because we lack sophisticated theory that goes beyond qualitative hypotheses regarding the signs of associations. Since we use the data by Tutić et al. (2024) we expected small strengths of association and a rather poor goodness of fit and therefore use pessimistic priors for regressions coefficients and the intercept (normal distribution with mean 0 and standard deviation 1). Regarding the prior of the dispersion parameter phi, we felt ignorant and hence picked a uniform distribution with a large upper bound. The second stage results in posterior distributions for the intercept, the regression coefficients, and dispersion parameter phi. Since we conduct these regressions to identify covariates for the DP-parameters and test our hypotheses, we are mostly interested in the posteriors of the indicators of social class and thinking dispositions.

The analysis in this study were conducted using Python and in particular the PyMC library. Data as well as code are available on OSF.[4]

Empirical results

Observed judgment in moral dilemmas

Before we summarize our findings regarding the estimates of the DP-parameters, we take a closer look at the empirical basis of these estimates. Recall that each respondent was confronted with five congruent and five incongruent dilemmas. Hence, the empirically observed judgment of each subject can be summarized by two numbers, p_con and p_incon, which give the proportions of congruent and incongruent dilemmas, respectively, in which this subject judged the described action as morally inappropriate. By design, both p_con and p_incon can only take six possible values, i.e., 0, 0.2, 0.4, 0.6, 0.8, and 1. Consequently, there are 36 possible patterns of individual

[4] https://osf.io/rc79e/?view_only=4f6e4373fbc44e20afbeedd898f2d9c4.

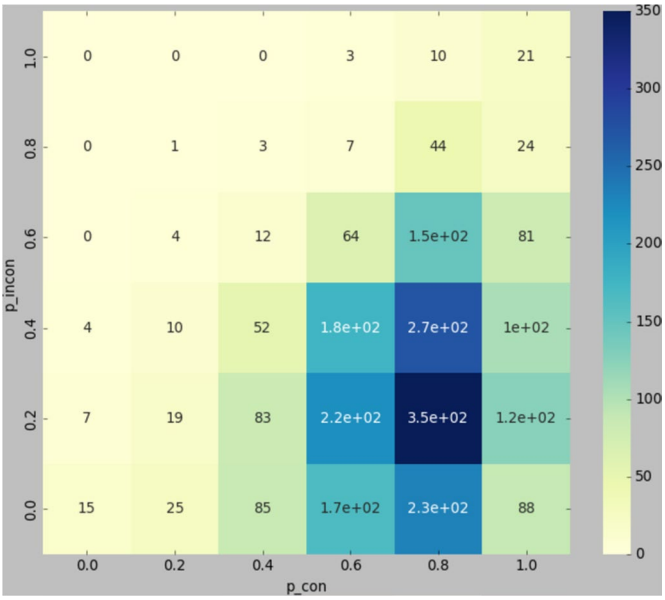

Fig. 5 Joint distribution of p_con and p_incon across individuals

judgment. Figure 5 provides an overview of the empirical observed joint distribution of p_con and p_incon across individuals. Empirically, from the 36 possible patterns, 31 have been observed. As expected, the great bulk of individuals are more judgmental in congruent than in incongruent dilemmas such that p_con > p_incon.

For each individual, our estimates of the DP-parameter are solely based on her pattern of moral judgment, i.e., her p_con and p_incon. So, in principle two subjects with identical observed judgment should be assigned identical DP-parameters. In practice, due to the inherent randomness in the MCMC estimation procedure, there are marginal variations in these estimates.

Estimating the dp-parameters

The left-hand side of Fig. 6 provides the posterior distributions of the as-, rc-, and epsilon-parameter averaged over all the individuals in our sample. The mean of as equals 0.27, the mean of rc equals 0.5, and the mean of epsilon equals 0.36. The model is very confident in these estimates as witnessed by the small HDIs, spanning from 0.27 to 028 (as), 0.49 to 0.5 (rc), and 0.36 to 0.37 (epsilon) respectively. This preciseness in the estimates of the mean parameters is due to the relatively high number of respondents in our sample. The shape of these posterior distributions is roughly normal; this is a direct consequence of the central limit theorem which implicates that the mean of a sufficient large number of distributions is approximately normal, regardless of the underlying distributions.

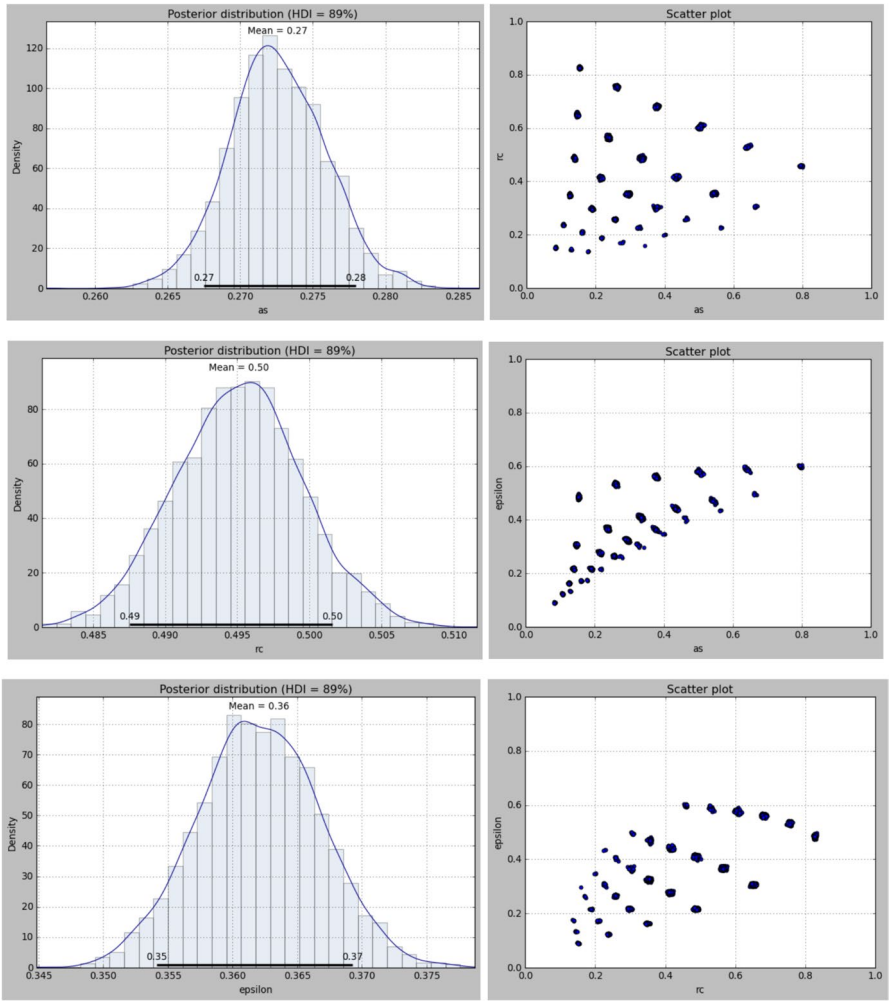

Fig. 6 Posterior distributions of the means of the DP-parameters as well as scatter plots of the parameters on the individual level

In fact, the posterior distributions on the individual level are far from normal. For illustrative purposes, Fig. 7 plots the posterior distributions of the DP-parameters for six respondents. Already from these six cases, we can learn that there is considerable variance in the means of the individual posteriors and a lot of uncertainty regarding the estimates on the individual level. For instance, the HDI for the as-parameter of subject 1 spans the interval from 0.06 to 0.73. While we have a rather large amount of data across individuals and therefore very precise estimates on the aggregate level, on the individual level there are just ten decisions from which we estimate three parameters. In addition, each behavioral pattern can be explained by multitude of configurations of DP-parameters. Both the scarcity of data on the individual level

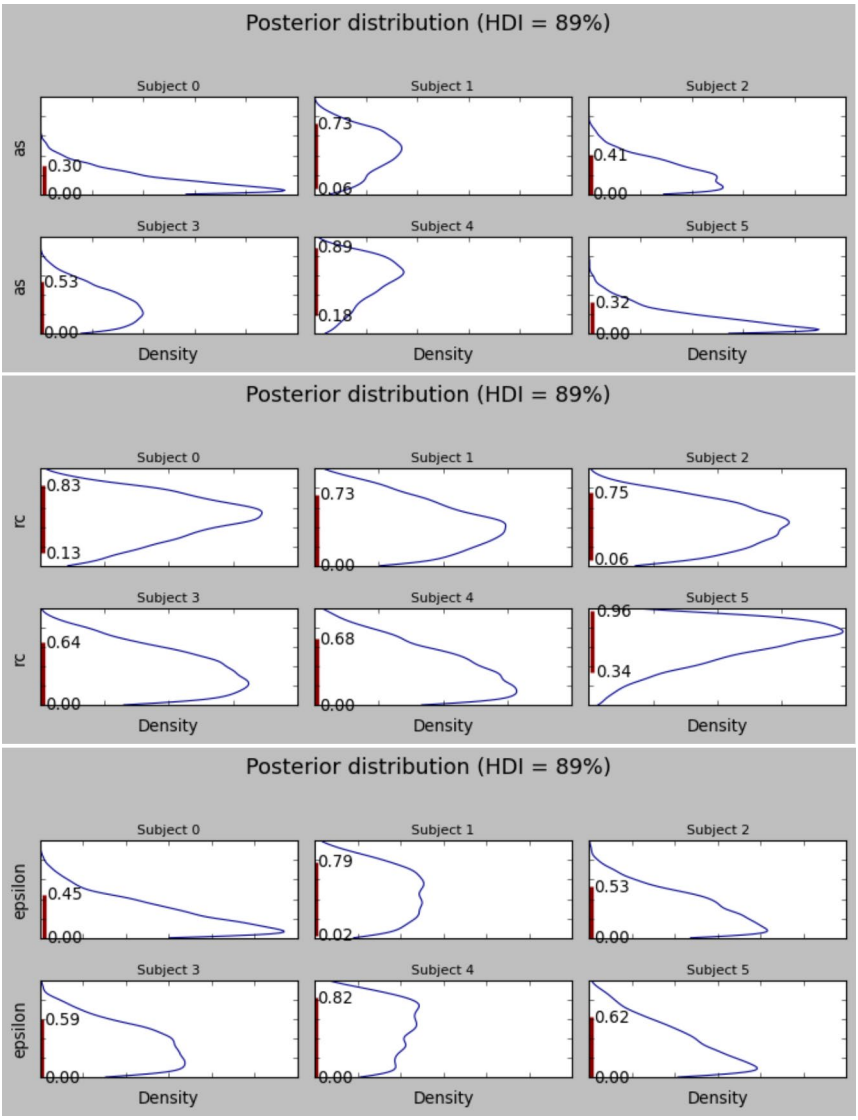

Fig. 7 Posterior distributions of the DP-parameters of six individuals

and the richness of possible explanations of behavioral patterns, contribute towards the high insecurity in our estimates on the individual level.

We will use the mean of the individual posteriors distributions as the estimates of the respective DP-parameter on the individual level. In addition, extended models in the supplementary materials will also make use of the standard deviations of the individual posteriors to account for the insecurity in these estimates. Table 1

provides an overview over the descriptive statistics of both means and standard deviations over the whole sample.

The three scatter plots of the DP-parameters (right-hand side of Fig. 7) highlight other important features of our Bayesian estimates. First, each scatter plot consists of 31 distinct points, which correspond to the 31 observed patterns of moral judgment, plus some "jitter" which stems from the inherent randomness of MCMC estimation. Second, the scatter plots of the parameter estimates reveal considerable associations among them: There is a slight negative correlation (r = -0.09) between the estimates of the as- and the rc-parameters, a strong positive correlation (r=0.72) between the as- and the epsilon-parameter, and a moderate positive correlation between the rc- and the epsilon-parameter (r=0.58). Understanding the reasons underlying these associations is tough because a lot of factors are into play. First, our assumption of uniform priors on the DP-parameters together with the intricacies of the DP-model determine how a particular pattern of individual behavior is explained in terms of estimates of the DP-parameters. Second, the empirical distribution of patterns of individual behavior play into the observed correlations, i.e., the "thickness" of the points in the scatter plots.

Figure 8 can help in intuiting the observed correlations between our estimates of the DP-parameters by providing an overview of the estimated DP-parameters for each combination of p_con and p_incon. Since both as and epsilon contribute positively towards both toward p_con and p_incon, we get monotonically increasing estimates for as and epsilon in both dimensions, i.e., with increasing p_con and p_incon. This pattern already explains the strong positive correlation between as and epsilon across the individuals. The estimates for the rc-parameter increases in p_con for a given p_incon, and decreases in p_incon for a fixed p_con. This follows from the fact that the rc-parameter controls the difference p_con – p_incon, which, ceteris paribus, increases in p_con and decreases in p_incon. Now we turn to the more difficult question of how to explain the weak negative correlation between as and rc and the moderate positive correlation between rc and epsilon, while at the same time as and epsilon are positively correlated. Figure 8 reveals that as is more strongly affected by p_incon than epsilon, while epsilon is more strongly affected by p_con than as. The overall correlation between as and rc on the one hand and epsilon and rc on the other hand combines the horizontal aspect (constant p_incon, variable p_con), which increases the respective correlations, and the vertical aspect (variable p_incon, constant p_con), which decreases the respective correlation. Since the

Table 1 Descriptive statistics of the estimates of the DP-parameters

	mean	std.	min	max
as (mean)	0.27	0.13	0.08	0.80
as (std.)	0.16	0.03	0.07	0.23
rc (mean)	0.50	0.15	0.14	0.83
rc (std.)	0.22	0.03	0.12	0.30
epsilon (mean)	0.36	0.12	0.09	0.61
epsilon (std.)	0.22	0.04	0.08	0.29

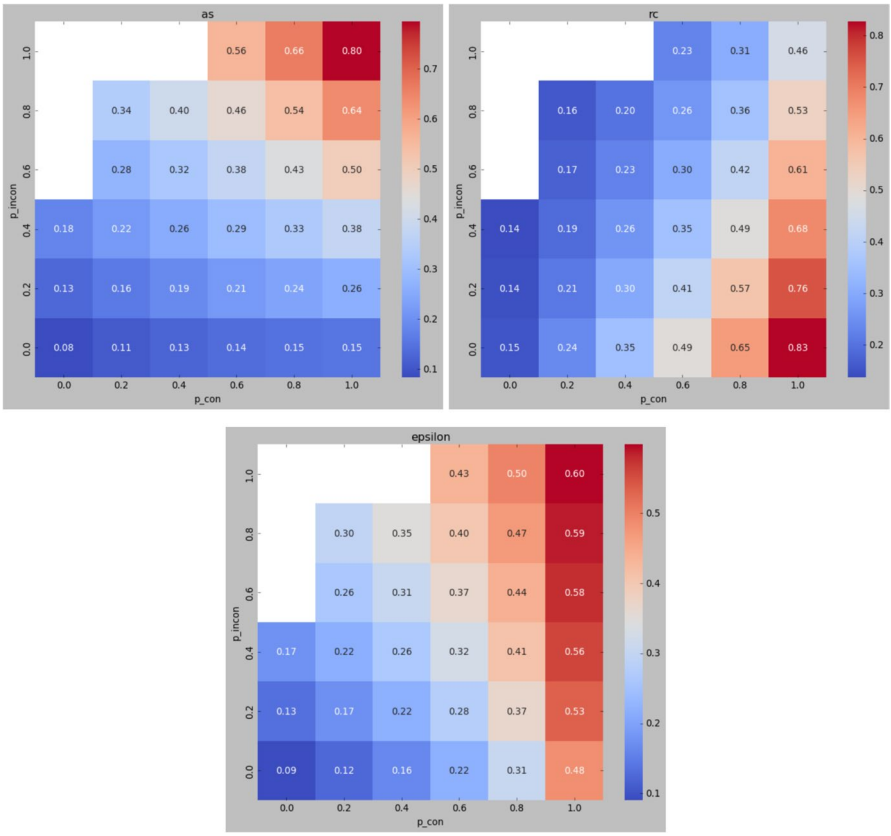

Fig. 8 Heatmaps showing the estimates of the DP-parameters as a function of observed p_con and p_incon

positive horizontal aspect is stronger with respect to epsilon than as and the negative vertical aspect is weaker, it makes sense that as and epsilon are differently associated with rc.

After describing our estimates of the DP-parameters and clarifying important basic relationships among them, we are now addressing the question of how well these estimates actually fit to the underlying empirical observations in the moral dilemmas. Figure 9 plots the empirically observed relative frequencies of judging the action harmful in both types of dilemmas (i.e., the aforementioned 31 behavioral patterns) as well as the relative frequencies predicted by the model. The corresponding points are connected by a line, respectively. The plot reveals several interesting features of the model: First of all, the model does generally not match the empirical proportions. The reason for that is tied to how Bayesian learning works: The empirical proportions are maximum likelihood estimates and, in this sense, appealing; but since we started with a uniform prior on as, rc, and epsilon we also started with an implicit prior on p_con and p_incon. According to this prior, the expected mean of

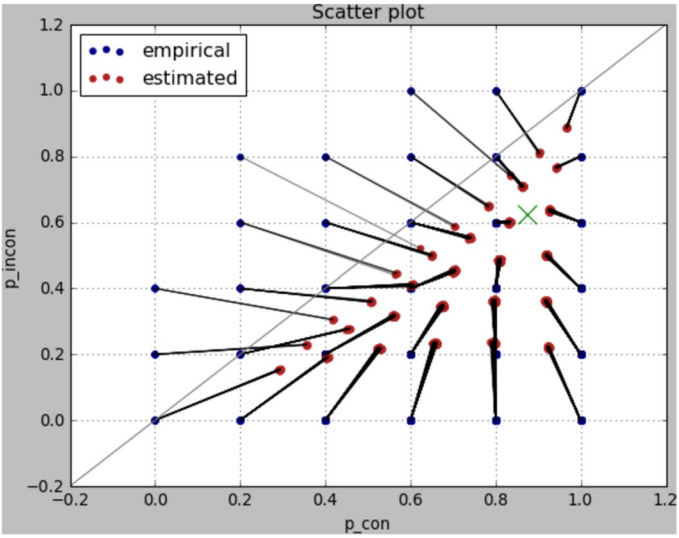

Fig. 9 Connected scatter plots of the empirically observed and retrodicted p_con and p_incon

p_con is 0.875 and the expected mean of p_incon is 0.625; this point is given by the green cross in Fig. 9.[5] This prior gets updated into the posterior by being confronted with the data, but only gradually. So, our estimates regarding these two probabilities are compromises between our prior and the observed relative frequencies. Since we have only limited data on the individual level, the weight of the data in this compromise is restricted.

Second, the model very decisively deviates from all empirical observations which have a higher probability of judging the action as harmful in incongruent than in congruent dilemmas. Of course, the reason for that is such a pattern can only be explained by the model as an outlier chance event but not structurally, since (1-as)rc cannot be negative. Third, the model is skeptical towards extremes, i.e., probabilities of 0 and 1. This is a direct consequence of using uniform priors which put considerable probability mass on non-boundary, interior points in the probability space.

In sum: The retrodictions based on the estimated DP-parameters do deviate substantially from the empirical observations on which these estimates are based. This is inherent in Bayesian estimation which next to empirical observations also factors in prior assumptions. The stronger an observation deviates from the logic of the estimated dual-process model or from the joint prior, the greater the deviation. These deviations should be considered as a feature, not a bug: Empirical data themselves are but imperfect realizations of a data-generating process mixed with randomness.

[5] We work with uniform priors on the unit interval for all three parameters; these have mean 0.5. According to the model, p_con = as + (1-as)rc + (1-as) (1-as)epsilon. Hence, p_con has the mean 0.5 + (1-0.5)0.5 + (1-0.5)(1-0.5)0.5 = 0.875 under the prior. A similar calculation obtains for p_incon.

We aim for recovering the underlying structural model and get rid of the randomness in data, which is why overfitting must be avoided.

Identifying covariates of the DP-parameters

In this section we use beta regressions with a logistic link function to identify covariates of our estimated DP-parameters. We control for various sociodemographic characteristics and focus on two indicators of social class, i.e., education and income, as well as two indicators of thinking dispositions, i.e., CRT-score and FI-score. In this subsection, we focus on the as- and the rc-parameter and, for simplicity, estimate models which do not take into account the insecurity in these estimates. The reader is referred to the supplementary materials which contains both models with epsilon as the dependent variable as well as models which include the measurement error in the DP-parameters; substantially, the extended models with measurement error do not lead to different results than the simpler models presented here.

as-Estimates. Figure 10 presents the results for the first model which uses the as-estimates as dependent variable. The coefficients depicted in the left panel are on the logodds-scale and therefore hard to interpret above and beyond their sign; to intuit effect sizes of variables of interest we will therefore make use of counterfactual plots in a second step. Starting with the variables of interest, we find that the great bulk of the probability mass in the posterior distribution of income lies to the left of 0; hence there is evidence that income is negatively associated with the as-estimates. Other than that, the model provides no clear evidence regarding associations between variables of interest and the as-estimates. Regarding the controls, we can be confident in concluding that females are more inclined towards making moral judgment in the as-mode than males and that people living in the eastern part of Germany are less inclined to the as-mode than people living in the western part. In addition to the regression coefficients, there are just two parameters estimated by this model: The intercept (mean = -1.06, 89%-HDI = [-1.13,-0.99]) and the concentration of the beta distribution (mean = 13.30, 89%-HDI = [12.62,13.99]).

The panel on the right side of Fig. 10 plots our estimates of the DP-parameters from the first stage against the posterior predictions of these estimates based on the

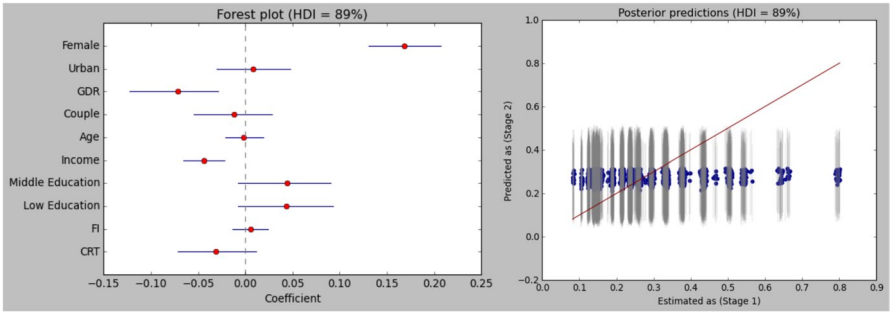

Fig. 10 Estimated coefficients (left panel) and posterior predictions (right panel) from a beta regression of as on social class and thinking dispositions (+controls)

beta regression. These predictions are not only based on the central tendency of the beta distribution, which we model via the covariates of the regression, but include also the randomness of this distribution as well. That is to say: For each observation, the blue point gives the average prediction, where we average over the whole joint posterior and over the samples of the beta regression. Hence, these predictions include both uncertainty in parameter values of the beta regression as well uncertainty in the sampling process. Apparently, the model does a bad job in predicting our as-estimates; basically, the model predicts values close to the overall mean of the as-estimates (0.27) regardless of the covariates. For low estimates of the as-parameter up to approx. 0.5 these predictions do hit the mark, i.e., the red line with slope 1 cuts across the respective HDI, but for higher values of the as-estimates, the predictions of the beta regression are far off. From this we can conclude that the variance in the as-parameter between our respondents can hardly be explained by social class, thinking dispositions, or the controls.

We did find however some evidence that income is negatively associated with the as-estimates. Figure 11 depicts the strength of this association by showing counterfactual predictions regarding the as-estimates with varying income, where the other covariates are either fixed at their empirical means (metric variables) or their mode (dummies). The black line indicates the average prediction regarding the central tendency of the beta distribution, the red HDI the 89% most credible values of this central tendency, and the grey HDI the 89% most credible predictions that include the randomness of the beta distribution. Put differently: The red HDI depicts parametric uncertainty, whereas the grey HDI include both parametric as well as sampling uncertainty. From the figure it is apparent that the association between income and the tendency to judge in the as-mode is rather weak. Numerically, the poorest actors are predicted to have an as-estimate of 0.30 and the richest actors are predicted to have an as-estimate of 0.25.

rc-Estimates. Turning to the beta regression which uses the rc-estimates as dependent variable, we present Fig. 12 which is analogous to Fig. 10 (additional

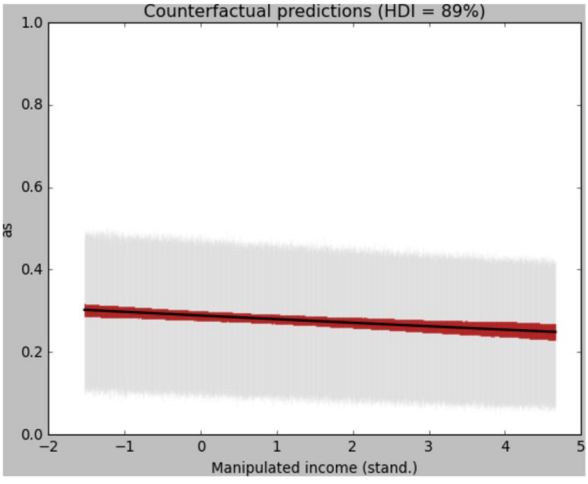

Fig. 11 Counterfactual predictions of as-estimates based on income

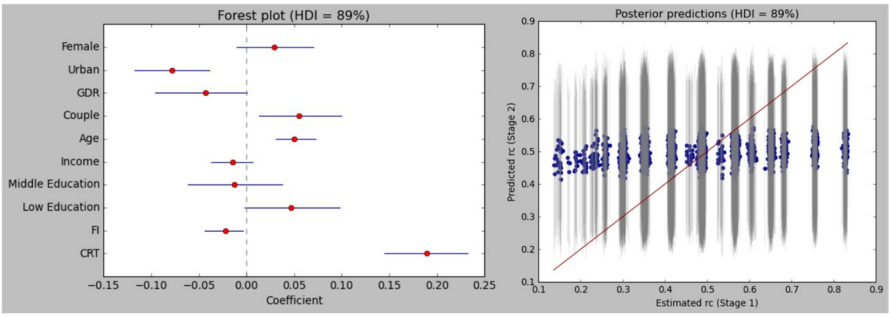

Fig. 12 Estimated coefficients (left panel) and posterior predictions (right panel) from a beta regression of rc on social class and thinking dispositions (+controls)

parameter estimates: Intercept (mean = -0.10, 89%-HDI = [-0.16,-0.02]); concentration of beta distribution (mean = 10.11, 89%-HDI = [9.60,10.63]). Interestingly, a rather different picture emerges regarding the posterior distributions of the covariates than in the as-case. The model provides convincing evidence that both indicators of thinking dispositions are associated with the tendency to make judgments in the rc-mode. That is, the better an actor fares in the CRT-test and the lower the actor scores on the FI-scale, the greater her predicted rc-estimate. In particular, the CRT-score seems to be an important predictor of the inclination towards moral judgment in the rc-mode. While there is no evidence that income is associated with rc-estimates or that actors with higher education differ from the actors with middle education, the model provides some weak evidence that actors with low educations might have a greater rc-estimates than actors with high education. Regarding the controls, we note that actors living outside of cities, actors living in the western part of Germany, actors in stable relationships, and older actors exhibit a greater inclination towards judgment in the rc-mode than their respective counterparts.

Regarding model fit, the beta regression using the rc-estimates as the dependent variable does not really do better than the one using the as-estimates. There is more variation in the predictions but little covariation with the rc-estimates. Notice the insecurity in the predictions which span a majority of the theoretically possible values, i.e., approx. the interval from 0.2 to 0.8. From these observations we conclude that our rc-estimates are hard to explain with the covariates at hand.

Turning to an evaluation of strengths of associations, we present Fig. 13 which depicts counterfactual plots for variables of interest. Regarding the FI-score, the prediction for actors with lowest FI-score equal 0.52 and for actors with highest FI-score 0.48. Actors with a high CRT-score are predicted to have a rc-estimate of 0.54 and actors with a low CRT-score are predicted to have a rc-estimate of 0.50. Regarding education, predictions between actors with high education (0.49) and low education (0.50) differ only minimally. In sum then we find weak associations between rc-estimates and CRT and FI, whereas education and rc-estimates are basically unrelated.

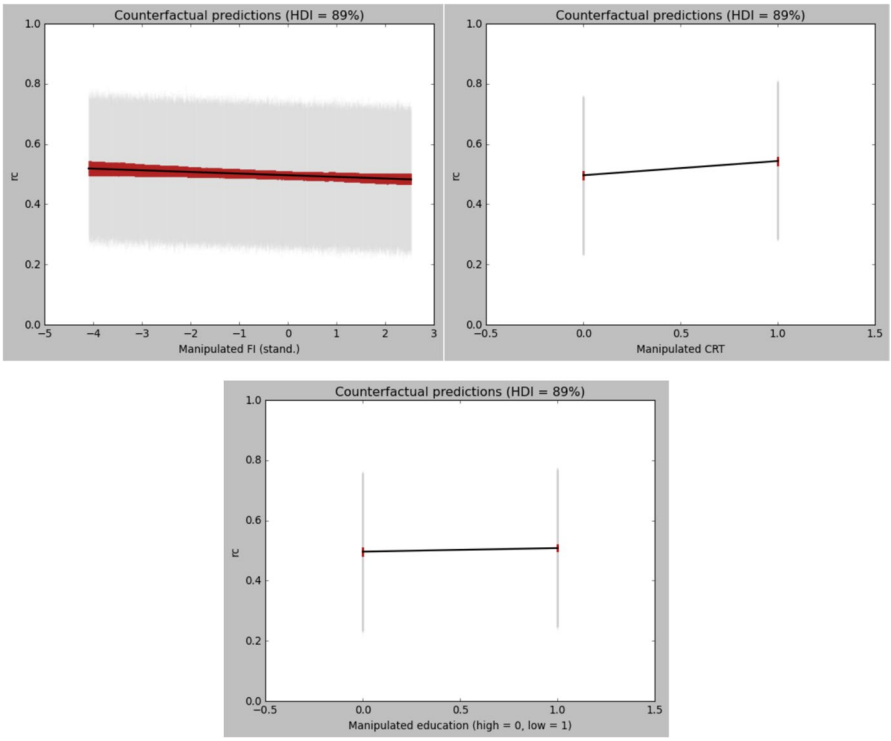

Fig. 13 Counterfactual predictions of rc-estimates based on FI-score, CRT-score, and education

Mediation

To check whether or not the associations between class and the DP-parameters are mediated by thinking dispositions, we can compare the two models above (full models; FM) with the respective models in which FI and CRT are not included as predictors (restricted models; RM). For a more elaborate discussion of mediation, the reader is referred to the supplementary materials.

Figure 14 provides the estimates of the coefficients for the class indicators for the two models with the as-parameter as the dependent variable. While the means of the posteriors of all coefficients are in fact slightly farer away from zero in the restricted than in the full model, the differences are very small and the credible intervals are very much overlapping. Hence, we find no evidence for the hypotheses that class-based differences in the inclination towards deontological judgment are mediated by thinking dispositions. Given our previous finding that thinking dispositions are not associated with the as-parameter (see Fig. 10), this result was expected.

Regarding the rc-parameter, Fig. 15 presents the posteriors of the relevant coefficients for both the full and the restricted model. If we focus on income and the comparison of individuals with low versus high education, we find that the coefficients in the restricted model are actually closer to zero than in the full model. Since higher-class individuals tend towards having more reflective

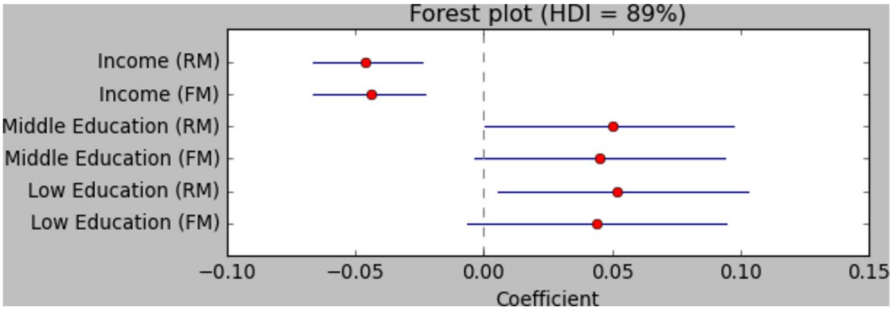

Fig. 14 Estimated coefficients of class indicators in models including (FM) and excluding (RM) indicators of thinking dispositions (+controls)

thinking dispositions (see supplementary material for details) and thinking dispositions are positively associated with the rc-estimates (see Fig. 12), this suggests that the positive mechanism which links social class to rc-estimates via thinking dispositions is counterbalanced by some unidentified other mechanisms contributing towards a negative association. While these differences in sizes of coefficients between the full and restricted model are interesting, none of these differences are big enough considering the insecurity in these estimates to warrant the conclusion that class-based differences in rc-estimates are mediated by thinking dispositions.

Discussion

In this paper, we used Bayesian inference to estimate a dual-process model in an attempt to study class-based differences in moral judgment. The model employs three parameters—deontological inclination (as), consequentialist inclination (rc), and judgmental tendency (epsilon)—to explain an individual's observed moral judgment. We found a mean as of 0.27, rc of 0.5, and epsilon of 0.36, with

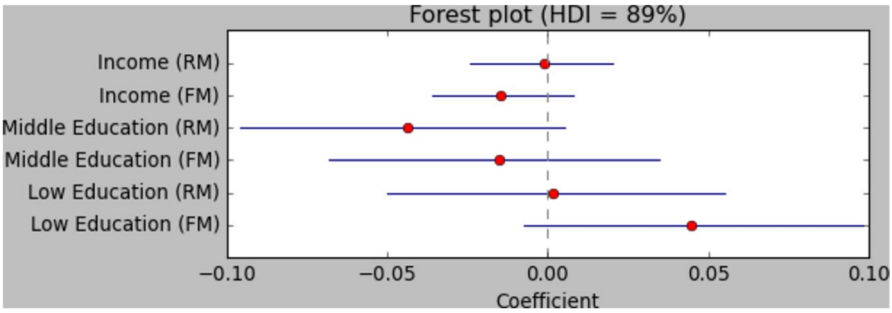

Fig. 15 Estimated coefficients of class indicators in models including (FM) and excluding (RM) indicators of thinking dispositions (+controls)

considerable insecurity in these estimates due to data scarcity. Using Bayesian beta regressions, we examined how social class and cognitive dispositions (measured by FI and CRT scores) are associated with these parameters. Our results show that while social class (specifically income) is associated with the as-parameter, the rc-parameter and social class are unrelated. Conversely, cognitive dispositions are only correlated with the rc-estimates, but not associated with the as-parameter. Finally, our study does not provide evidence for the hypothesis that the relationships between social class and utilitarian or deontological inclinations are mediated by thinking dispositions.

Comparing our findings to existing literature on social class and moral judgment, i.e. Côté et al. (2013) and Tutić et al. (2024), our study cum grano salis supports the latter's main finding in that social class is more strongly associated with deontological than with utilitarian inclinations. At the same time, there is also the discrepancy that thinking dispositions in the former study do mediate the relationship between class and utilitarianism whereas the present study does not corroborate this claim. In interpreting these findings, it has to be kept in mind that while this paper employs a different dual-process model of moral judgment and a Bayesian estimation strategy, it nevertheless uses the same data as Tutić et al. (2024), and therefore cannot be seen as fully independent evidence against the findings of Côté et al. (2013).

The findings of this study contribute to a nuanced understanding of how moral judgment varies across social classes by drawing on three theoretical perspectives: contextualism and solipsism, interaction ritual theory, and cognitive styles. Each of these approaches, despite their distinct mechanisms, can be interpreted to suggest a positive relationship between social class and utilitarian judgment and a negative correlation between social class and deontological judgment. For instance, the contextualism and solipsism framework explains class-based differences in terms of material constraints, with lower-class individuals exhibiting more communal orientations and higher-class individuals demonstrating solipsistic, agentic behavior. Similarly, interaction ritual theory suggests that the social diversity of higher-class individuals leads to more open-minded, universalistic, and utilitarian forms of moral reasoning, whereas lower-class actors rely on particularistic and taken-for-granted deontological judgments. However, our data only provide partial support for these claims: While we do observe that higher social class—especially income—correlates negatively with deontological judgment, we find no significant positive correlation between social class and utilitarianism. While speculative, these findings suggest that the relationship between social class and moral judgment is considerably more complex than previously acknowledged in the literature (Côté et al., 2013; Tutić et al., 2024). That is, while higher-class individuals may be less prone to defaulting to deontological snap judgments, their reflective capacities do not necessarily translate into utilitarian reasoning. Instead, they may engage in more context-sensitive forms of moral judgment that draw on a broader range of considerations, such as social norms, reputation management, or moral identity (Aquino & Reed, 2002). For example, higher-class individuals might consider the potential social ramifications of their actions or the ways in which their decisions align with their moral identity, rather than adhering strictly

to utilitarian principles. This flexibility in moral reasoning may explain why the expected strong correlation between class and utilitarian judgment did not emerge. Clearly, further empirical research is needed to explore how these cognitive and contextual factors interact across different social strata to shape moral judgments more fully.

Regarding the third theoretical approach considered in this paper—namely, dual-process research on the role of cognitive styles—this study provides additional insights beyond the observed associations between social class, utilitarianism, and deontology. Specifically, it sheds light on the relationships between social class and cognitive styles, the connections between cognitive styles and moral judgment, and the potential mediating role of cognitive styles in the class-judgment relationship. From both an action-theoretical perspective and within the framework of the dual-process model of moral judgment (Greene, 2007, 2013), the associations between cognitive styles and the as- and the rc-parameters are particularly relevant. Consistent with previous literature, we found that thinking dispositions are closely associated with the rc-parameter. Although the associations between thinking dispositions and the as- parameter are weaker, at least descriptively, they align with theoretical expectations. In fact, only 12% of the posterior probability mass for the coefficient of the CRT score lies above 0, while 88% lies below zero. Overall, these findings support the idea that cognitive styles are related to both utilitarian and deontological judgment, as suggested by Greene's dual-process model. Furthermore, we found evidence that higher social classes exhibit a greater tendency toward cognitive reflection than lower social classes (Brett & Miles, 2021). However, contrary to our expectations, these links between class and cognitive styles, and between cognitive styles and moral judgment, do not significantly explain class-based differences in moral judgment, as there is no substantial evidence for cognitive styles as mediators. This suggests that while thinking dispositions play a role in the complex relationship between social class and moral judgment, other factors such as empathic concern or moral identity (Tutić et al., 2024) may also be crucial and warrant further exploration.

Regarding the limitations of the present study, we want to highlight two important aspects. First, as explained in the first part of the results section, the estimation of the DP-parameters on the individual level relied on barely ten observed judgments in moral dilemmas. While Bayesian estimation can work well with small samples, this scarcity of data on the individual level led to rather huge influence of prior distributions on the DP-estimates. Future research should therefore check whether these findings can be replicated with more data on the individual level. The second main limitation of the current study lies in the fact that it is observational. While objective indicators of social class such as income and education eschew experimental manipulation, subjective indicators such as actor's perception of her own status can be randomized, thereby facilitating the identification of causal effects. Similarly, there exist well-established techniques to induce intuitive or deliberative decision making and judgment (e.g. Ferreira et al., 2006; Tutić et al., 2022) which can and should be substituted for barely measuring thinking dispositions in future research.

All in all, Bayesian estimation proved to very useful in conducting theory-driven empirical research in this application. The method allowed us to estimate precisely the model we wanted to for theoretical reasons, closing the gap between theoretical and statistical modelling. In addition, the Bayesian focus on posterior distributions was helpful in staying aware of the insecurity in our estimates and hence cautionary in our substantial inferences. Against this background, we feel that the combination of Bayesian estimation with the explicit modelling of theoretically informed decision procedures is a promising avenue for future research in sociological action theory.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11186-024-09584-1.

Funding Open access funding provided by University of Bergen (incl Haukeland University Hospital).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57–86.

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Bourdieu, P. (1987). *Distinction. A social critique of the judgment of taste*. Harvard University Press.

Brett, G., & Miles, A. (2021). Who thinks how? Social patterns in reliance on automatic and deliberate cognition. *Sociological Science, 8*, 96–118.

Ciaramelli, E., Muccioli, M., Làdavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2*, 84–92.

Clark, M. S., & Mills, J. (1993). The difference between communal and exchange relationships: What it is and is not. *Personality and Social Psychology Bulletin,19*, 684–691.

Collins, R. (1988). *Theoretical sociology*. Harcourt Brace Jovanovich.

Collins, R. (2004). *Interaction ritual chains*. Princeton University Press.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*, 216–235.

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition, 179*, 241–265.

Côté, S., Piff, P. K., & Willer, R. (2013). For whom dothe ends justify the means? Social class and utilitarian moral judgment. *Journal of Personality and Social Psychology, 104*, 490–503.
Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D.C: Heath.
Durkheim, É. [1915]. (2008). *The elementary forms of the religious life*. Dover.
Ekeh, P. P. (1974). *Social exchange theory: The two traditions*. Harvard University Press.
Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71*, 390–405.
Esser, H., & Kroneberg, C. (2015). An integrative theory of action: The model of frame selection. In E. J. Lawler, S. R. Thye, & J. Yoon (Eds.), *Order on the Edge of Chaos: Social Psychology and the Problem of Social Order* (pp. 63–85). Cambridge University Press.
Evans, J. S. B. T. (2010). *Thinking twice. Two minds in one brain*. Oxford University Press.
Evans, J. S. B. T. (2018). Dual process theory: Perspectives and problems. In De W. Neys (Ed.), *Dual process theory 2.0* (pp. 137–155). Routledge.
Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science,8*, 223–241.
Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology, 91*, 797–813.
Fleischmann, A., Lammers, J., Conway, P., & Galinsky, A. D. (2019). Paradoxical effects of power on moral thinking: Why power both increases and decreases deontological and utilitarian moral decisions. *Social Psychological and Personality Science, 10*, 110–120.
Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review,5*, 5–15.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives,19*, 25–42.
Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology,66*, 8–38.
Goffman, E. (1967). *Interaction Ritual: Essays on face to face Behavior*. Doubleday/Anchor.
Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences,11*, 322–323.
Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.
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*, 1144–1154.
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*, 389–400.
Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review,108*, 814–834.
Haidt, J. (2003). The Moral Emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 852–70). Oxford University Press.
Hooghe, M., Reeskens, T., Stolle, D., & Trappers, A. (2009). Ethnic diversity and generalized trust in Europe: A cross-national multilevel study. *Comparative Political Studies,42*, 198–223.
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language,30*, 513–541.
Kahneman, D. (2011). *Thinking, fast and slow*. Penguin Books.
Kant, I. (1797). Die Metaphysik der Sitten. In zwey Theilen. Teil 1: Metaphysische Anfangsgründe der Rechtslehre. Friedrich Nicolovius.
Kelley, C. M., & Jacoby, L. L. (2000). Recollection and familiarity: Process-dissociation. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 215–28). Oxford University Press.
Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446*, 908–911.
Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science,20*, 99–106.
Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., & Keltner, D. (2012). Social class, solipsism, and contextualism: How the rich are different from the poor. *Psychological Review,119*(3), 546–572.

Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie, 34*, 344–363.
Kroneberg, C. (2011). Die Erklärung sozialen Handelns. Grundlagen und Anwendung einer integrativen Theorie. VS Verlag für Sozialwissenschaften.
Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychological Review, 118*, 97–109.
Kruschke, J. K. (2015). *Doing bayesian data analysis*. Academic.
Leamer, E. E. (1983). Let's take the con out of Econometrics. *American Economic Review,73*, 31–43.
Lee, J. J., & Gino, F. (2015). Poker-faced morality: Concealing emotions leads to utilitarian decision making. *Organizational Behavior and Human Decision Processes,126*, 49–64.
Li, Z., Wu, X., Zhang, L., & Zhang, Z. (2017). Habitual cognitive reappraisal was negatively related to perceived immorality in the harm and fairness domains. *Frontiers in Psychology, 8*, 1805.
Lindenberg, S. (2008). Social rationality, semi-modularity and goal-framing: What is it all about? *Analyse & Kritik,30*, 669–687.
Lizardo, O. (2017). Improving cultural analysis: Considering personal culture in its declarative and non-declarative modes. *American Sociological Review,82*, 88–115.
Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science,341*, 976–980.
McElreath, R. (2020). *Statistical rethinking*. CRC.
Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences,22*, 280–293.
Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology, 18*, 193–197.
Miles, A. (2015). The (re)genesis of values: Examining the importance of values for action. *American Sociological Review,80*, 680–704.
Miles, A., Brett, G., Khan, S., & Samim, Y. (2023). Testing models of cognition and action using response conflict and multinomial processing tree models. *Sociological Science, 10*, 118–149.
Mill, J. S. (1863). *Utilitarianism*. Parker, Son and Bourn, West Strand.
Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition,84*, 221–236.
Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review,11*, 988–1010.
Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology, 20*, 272–314.
Payne, B. K., & Cameron, C. D. (2014). Dual-process theory from a process dissociation perspective. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-Process Theories of the Social Mind* (pp. 107–120). The Guilford Press.
Petrinovich, L., O'Neil, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology, 64*, 467–478.
Piazza, J., & Landy, J. F. (2013). Lean not on your own understanding': Belief that Morality is founded on Divine Authority and Non-utilitarian Moral judgments. *Judgment and Decision Making*, *8*, 639–661.
Pichler, F., & Wallace, C. (2009). Social capital and social class in Europe: The role of Social Networks in Social Stratification. *European Sociological Review, 25*, 319–332.
Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences,109*, 4086–4091.
Reynolds, C. J., & Conway, P. (2018). Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. *Emotion, 18*, 1009–1023.
Stanovich, K. E. (Ed.). (2011). *Rationality and the reflective mind*. Oxford University Press.
Thomson, J. J. (1986). *Rights, restitution, and risk: Essays in moral theory*. Harvard University Press.
Tutić, A., Haiser, F., & Krumpal, I. (2024). Social class and moral judgment: A process dissociation perspective. *Frontiers in Sociology, 9*, 1391214.
Tutić, A., Krumpal, I., & Haiser, F. (2022). Triage in times of COVID-19: A moral dilemma. *Journal of Health and Social Behavior, 63*, 560–576.
Tutić, A. (2022). Cultural orientations and their influence on social behaviour: Catalysation and suppression. *Journal for the Theory of Social Behaviour, 52*, 438–453.

Tutić, A., & Liebe, U. (2020). Contact heterogeneity as a mediator of the relationship between Social Class and Altruistic giving. *Socius: Sociological Research for a Dynamic World*. https://doi.org/10.1177/2378023120969330

Tutić, A., & Liebe, U. (2019). Sozialer Status, Altruistisches Geben und Reziprozität: Befunde aus einem Quasi-Experiment mit Probanden aus den USA. *Zeitschrift für Soziologie,48*, 176–189.

Uehara, E. (1990). Dual exchange theory, social networks, and informal social support. *American Journal of Sociology,96*, 521–557.

Vaisey, S. (2009). Motivation and justification: A dual-process model of culture in action. *American Journal of Sociology,114*, 1675–1715.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language,46*, 441–517.

Zhang, L., Li, Z., Wu, X., & Zhang, Z. (2017). Why people with more emotion regulation difficulties made a more deontological judgment: The role of deontological inclinations. *Frontiers in Psychology,8*, 2095.

Andreas Tutić is Professor of Sociology at the University of Bergen, Norway. His research focusses in sociological action theory and cognitive sociology, frequently employing modeling, simulation, and experimental methods to explore complex social phenomena. His work covers diverse topics such as moral judgment, prosocial behavior, and the interplay of social class and status. Tutić's scholarly contributions have been published in esteemed journals including Sociological Science, European Sociological Review, and the Journal of Health and Social Behavior.