Can a Distraction Task Reveal Hidden Ethnic Bias in Economics Games?*

Jennifer Hamilton[1], Chad Hazlett[1,2], and Daniel N. Posner[1]

[1]Department of Political Science, UCLA
[2]Department of Statistics and Data Science, UCLA

September 3, 2024

Prepared for presentation at the 2024 American Political Science Association Annual Meeting

Abstract

Behavioral economics games such as the Dictator Game (DG) have been widely used to quantify ethnic and other between-group biases. However, in numerous contexts where ethnicity is thought to be extremely salient, the DG has failed to detect any evidence of ethnic bias. A compelling hypothesis for this failure is that economics games like the DG permit participants to alter their behavior to fit social norms that discourage ethnic bias or discrimination. Drawing on dual process models from psychology, we induce cognitive load through a concurrently played distraction task (the Spatial Delayed Recall Task) in a within-subjects experimental design. If self-monitoring explains the apparent lack of ethnic bias in the DG, and if self-monitoring can be reduced through such distraction, then we a proportion of bias may be "unmasked" when the DG is played concurrently with the distraction task. The difficulty of the SDRT was calibrated in a separate pilot study with over 200 participants. We deploy this design among 558 Kikuyu and Luo participants in Nairobi, Kenya, a setting in which previous DG studies failed to produce evidence of ethnic bias among these groups even though ample evidence can be found for the salience of such bias in daily life. Contrary to expectations, we find no evidence of differential ethnic bias across DGs conducted with and without the distraction task, notwithstanding evidence that the distraction task does in fact (modestly) increase cognitive load. We conclude that the inability of DGs to detect ethnic bias is likely not because they permit self-monitoring and social desirability bias.

*The authors thank Lornah Wahome, Cecile Juma, Joseph Kimani, Kelvin Kihindas, Alice Wambui, Patricia Omedo, Brian Baraza, and Jacob Posner for their indispensable help with project implementation. The study was funded with support from the UCLA Bedari Kindness Institute and the UCLA Committee on Academic Research. Human subjects approval was received from both Maseno University (MUERC/01116/22) and UCLA (IRB#22-001029). A pre-analysis plan was registered at osf.io.

Introduction

Discrimination and bias across group lines is a pressing concern in many societies, and both a source and indicator of social conflict. However, if discrimination and bias is to be addressed scientifically, it must first be measured. Our research identifies a well-known problem with a commonly used tool for measuring inter-group biases and proposes, and tests a novel solution.

The particular measurement tool we study is the Dictator Game (DG), an economics game that has been widely used by social scientists to quantify ethnic and other between-group biases.[1] Yet, there are reasons to suspect that the DG (and other behavioral economics games like it) may generate inaccurate or incomplete measures of inter-group bias. It has been well demonstrated that, despite the real-money stakes used to incentivize behavior, the DG may be influenced by social expectations, including the expectation that one should not reveal one's biases toward members of other groups (Haley and Fessler, 2005; List, 2007; Levitt and List, 2007; Bardsley, 2008; Zizzo, 2010; Cilliers, Dube and Siddiqi, 2015; de Quidt, Haushofer and Roth, 2018). This susceptibility stems from a combination of participants' knowledge that their decisions are being scrutinized and the opportunity provided in the DG for participants to reflect before making their offers, and thus control their responses. In contexts where participants expect open displays of cross-group bias to be frowned upon, they may conceal their biases, generating outcomes in the game that are not reflective of their underlying biases toward out-group members. The resulting measures of bias that DGs generate may thus lead to inaccurate conclusions and faulty theorizing.

The possibility that economics games may fail to pick up biases that actually exist is consistent with the robust and well-powered null results for ethnic bias in several studies that use such games in locations where ethnic biases are thought to be strongly present. For example, Berge et al. (2020), Blum, Hazlett and Posner (2021), and Barriga et al. (2023) all find no evidence of ethnic bias in either the DG or the Public Goods Game (PGG) when played by large samples of residents of Nairobi, Kenya.[2] Habyarimana et al. (2009) also find no evidence of ethnic bias in either the standard DG or PGG played in Kampala, Uganda—a setting where ethnicity is similarly important in political and social interactions. Pilot data for this project, as well as the DG results presented below, replicate these null findings.[3] By contrast, the same participants who showed no evidence of ethnic bias in their

[1]We provide a complete description of the game in the discussion of our protocol below. A partial list of studies using the DG to measure cross-group bias includes Fershtman and Gneezy (2001); Whitt and Wilson (2007); van der Merwe and Burns (2008); Habyarimana et al. (2009); Chen and Li (2009); Ahmed (2010); Adida, Laitin and Valfort (2014); McCauley (2014); Mironova and Whitt (2014); Iyengar and Westwood (2015); Peterson (2016); Tanaka and Camerer (2016); Jeon, Johnson and Robinson (2017); Berge et al. (2020); Blum, Hazlett and Posner (2021); Barriga et al. (2023); Haushofer et al. (2023), and Posner (2024).

[2]This includes (in both Berge et al. (2020) and Blum, Hazlett and Posner (2021)) experimental sessions taking place in the weeks leading up to national elections, when the salience of ethnicity is expected to be heightened (Eifert, Miguel and Posner, 2010).

[3]Haushofer et al. (2023) do, however, find evidence of ethnic bias in the DG in a Nairobi sample quite similar to the ones used in Berge et al. (2020), Blum, Hazlett and Posner (2021), and the present study.

DG behavior in Blum, Hazlett and Posner (2021) show the expected biases when implicit measures are instead used.

Economics games like the DG remain widely used tools, yet the inability of these games to detect the existence of ethnic biases in these cases, combined with the motivation and opportunity participants may have to hide their biases, raises questions about their utility as measurement tools and about the interpretation of prior results based on their output. It also challenges researchers to better understand the sources of this measurement failure.

Our Approach

We propose and implement a strategy for testing whether the inability of DGs to reliably measure ethnic biases stems from their vulnerability to social desirability bias and self-monitoring. Our strategy rests on the distinction made in dual-process theories, dating back to at least James (1890), between behavior guided by "automatic" reactions (sometimes referred to as "System 1") and more deliberative and reasoned control (variably referred to by names including executive control, effortful control, or "System 2"). Applied to the phenomenon of intergroup bias, these models propose that when a person perceives an out-group member, a range of semantic and affective associations are automatically stimulated, possibly outside of awareness, including associations driven by negative stereotypes (System 1, e.g. Payne, 2001). Left unchecked, these associations can lead to behaviors that exhibit out-group bias. But with time, attention, and effort (System 2) individuals are often able to override or conceal these impulses in their expressed behavior (see e.g., Devine (1989); Payne et al. (2005)).[4] Seen through this lens, the disconnect between the cross-group biases we observe in the world and the consistent null results in the cross-group DG (and also PGG) may stem from participants' successful efforts to employ cognitive resources (System 2) to bring their behavior into alignment with personal or social expectations. Our research introduces a means of short-circuiting participants' ability to do this by depleting the cognitive resources required by System 2.[5] This, we hypothesize, will mitigate participants' ability to successfully monitor and control their responses to satisfy social demands, resulting in more revealed ethnic bias in their DG allocations.

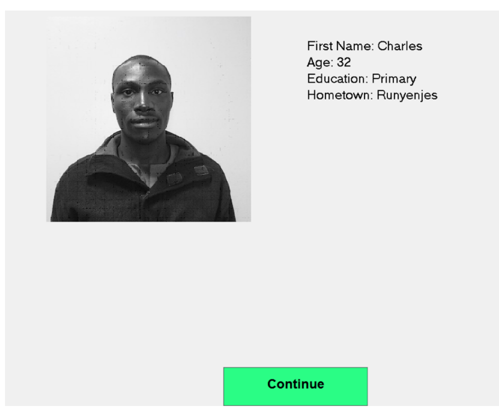
[4] Another line of reasoning, the Social Heuristics Hypothesis (see Rand et al., 2014 for review) posits that individuals may have a pre-potent impulse towards cooperation (though perhaps limited to members of social groups within a certain distance), but that in one-shot interactions like the DG, time to reflect can lead to more selfish responses than we would otherwise expect. We regard such processes as one of the influences on participants' preferred response *prior to* possible censorship due to social desirability bias. That is, the task design may lead to participants to have a "more selfish" preference because they have time to reflect on the task as they learn it, practice it, and play it with multiple partners. It is less clear how this would differ by the ethnicity of the partner. Nevertheless, the participants' "preferred" response given such processes, which we hope to measure, may be over-ridden by social desirability bias during game play. Our question is then whether we are able to better reveal that response through the manipulations we describe here.

[5] Our study is not the first to investigate the impact of cognitive load on behavior in the DG. For example, see Schulz et al. (2014) and Hauge et al. (2016).

Implementation of the DG

Before describing our strategy for reducing participants' ability to control their responses in the DG, we describe more precisely how the DG is played in our research. Participants ("dictators") are given an endowment of money (approximately 0.5 USD per round) and are asked to decide how much of this money to keep and how much to give away to a partner (the "receiver"). Whatever decision they make is implemented, with the money transferred to the receiver after the lab session. The receiver is not present in the room with the dictator, but rather is represented by a profile containing their picture, first name, age, education level, and hometown (as in Figure 1).[6] The key manipulation is provided by the receiver's hometown, which we experimentally vary across hometowns that, in the context we study, are immediately recognizable as being located in a Kikuyu or Luo part of the country, thus identifying the receiver as almost certainly Kikuyu or Luo.[7] To allow us to test for the impact of having no discernible ethnic cue, one third of the receivers' hometowns are identified as "Nairobi," whose multi-ethnic demography provides little information about the receiver's likely ethnic background. The additional information about age and education is included to obscure our interest in ethnic bias, as is standard in experiments of this type.[8]

Figure 1: Profile of Receiver Used in Dictator Game



[6]All of the pictures we use are of men, in order to reduce the number of factors that might affect participants' response to the profile. The photos were pre-screened to be ethnically ambiguous, so that the hometown cue would more likely shape the participants' perception of the receiver's ethnic group. Each participant saw each photograph and associated profile just once.

[7]Berge et al. (2020), who use the same hometown cues, provide evidence for their effectiveness as indicators of ethnic group affiliation in this context.

[8]Further information on the construction of the profiles, and on how we ensure that the experiment is deception-free, can be found in Blum, Hazlett and Posner (2021), whose protocols we adopted. These procedures also follow closely from prior protocols including those of Berge et al. (2020) and Habyarimana et al. (2009).

Short-circuiting System 2: the Spatial Delayed Recall Task (SDRT)

Our strategy for testing whether self-monitoring can explain the consistent null results found in prior work is to have participants play the DG while concurrently engaged in a distraction task that depletes the attentional resources they might otherwise apply to censoring biased responses. The task we employ, the Spatial Delayed Recall Task (SDRT), originates from Glahn et al. (2002) and has been used for a variety of research and clinical purposes including studies of working memory and ADD/ADHD in children (e.g. Lenartowicz et al., 2019; Hazlett et al., 2020).[9] An advantage of the SDRT for the present study is that it does not require skilled recall of letters or numbers and has been shown to work well in populations with a wide range of educational backgrounds and familiarity with computers.

In each trial, participants see either one or two "cue" dots in a variety of positions on the screen, followed by a blank screen (the "maintenance phase"), followed by a "probe" dot at a single location. Participants are instructed to indicate whether the probe dot was in the same location as the cue dot (or, in the case where there were originally two cue dots, in the same location as either of them). This requires that they maintain in working memory a representation of where the cue dot(s) was (were) located. The cognitive load (and difficulty of the task) is greater when there are more cue dots, so including both a one-dot (SDRT-1) and two-dot (SDRT-2) version allows us to test whether different levels of cognitive load have different effects. Figure 2 provides a schematic representation of the SDRT-2.[10] These difficulties levels were chosen after first testing a variety of difficulties on ourselves and lab members, followed by a pilot study that employed SDRT-1 and SDRT-3 combined with the DG as we do here. In the pilot phase, while the response to SDRT-1 was promising, SDRT-3 proved too difficult (responses were correct no more than chance). This prompted us to choose SDRT-1 and SDRT-2 for the full study described here.

Combining the DG and SDRT

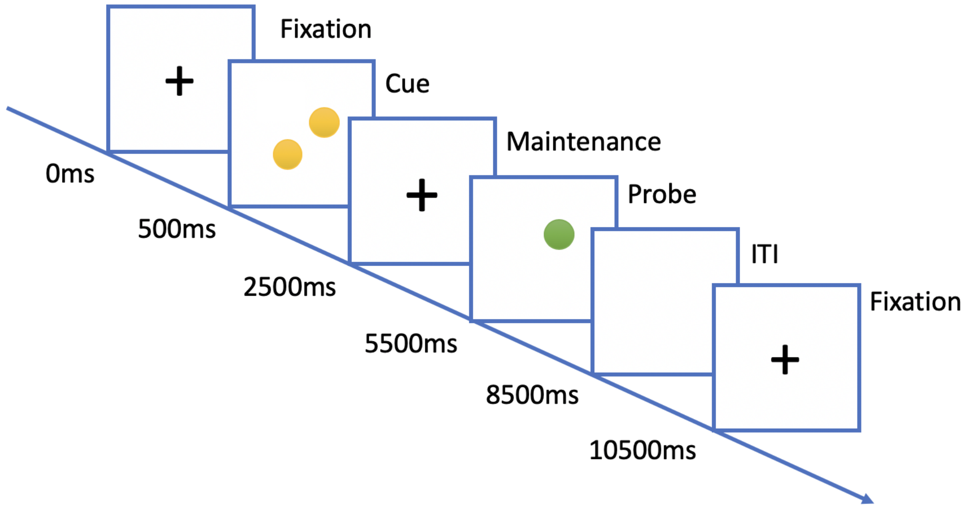
To learn how the distraction task affects allocations in the DG, we ask participants to play the DG *while* they perform the SDRT-1 or SDRT-2. We then compare the degree of ethnic bias we observe with rounds in which the participants play the traditional standalone DG, without distraction.

In the SDRT-1 and SDRT-2 treatments, participants first see the cue dot(s) of the SDRT, displayed for 2 seconds. Once the cue dot(s) disappear(s) (i.e., in the maintenance phase), a DG profile appears and participants choose how much to give to the receiver, just as they do in the standalone DG. Once they respond, the probe dot appears. The participant is then prompted to respond whether the probe

[9] We use cognitive load rather than time-pressure in our designs because speeded tasks rely on higher familiarity with computer-based tasks, and were found to be extremely frustrating for participants, who often disengaged when trials were presented too quickly.

[10] The timings, indicating total elapsed time in milliseconds (ms), reflect the ordinary use of this task in isolation. The timing for the maintenance phase is altered when we interrupt the SDRT with the DG, as described below.

Figure 2: SDRT Schematic



is in a location previously occupied by a cue dot.

Participants play the DG at all three levels of cognitive load (no distraction, SDRT-1, and SDRT-2) in randomized order. At each level of distraction, the DG is played three times: once with a coethnic, once with a non-coethnic, and once with a receiver with a neutral ethnic profile (i.e., from Nairobi), again in a randomized order. The degree of bias can be ascertained for each participant at each level of distraction. For example, the ethnic bias under the condition of no induced cognitive load ($bias_0$) is given by the contribution to the coethnic receiver minus the contribution to the non-coethnic receiver in the standalone DG. Ethnic bias under SDRT-1 ($bias_1$) and SDRT-2 ($bias_2$) (as well as pooling SDRT-1 and SDRT-2 together ($bias_{1,2}$)) are defined similarly as the difference between contributions to coethnics and non-coethnics under those levels of distraction.

Our expectation is that the attentional resources devoted to the SDRT will deplete the resources available for resisting the impulse to discriminate along ethnic lines, and thus that we will find higher ethnic bias in the DG when performed during the SDRT-1 and SDRT-2 than in the standalone DG. Our primary outcomes of interest are thus the contrasts across the mean allocations under these different treatments, averaged across all participants, all of which we expect to be positive:

$$\text{Distraction}_{1,2\text{vs}0} = \overline{bias_{1,2}} - \overline{bias_0}$$

$$\text{Distraction}_{1\text{vs}0} = \overline{bias_1} - \overline{bias_0}$$

$$\text{Distraction}_{2\text{vs}0} = \overline{bias_2} - \overline{bias_0}$$

$$\text{Distraction}_{2\text{vs}1} = \overline{bias_2} - \overline{bias_1}$$

Note that exposure to the distraction task need not fully eliminate social desirability bias for every participant or in every round. Our research strategy requires only that distraction is enough to unmask some proportion of bias, some proportion of the time, sufficient to measure a difference in the average contributions by level of distraction.

Data Collection and Experimental Protocols

We recruited 558 Kikuyu and Luo participants to participate in our study at the Busara Center for Behavioral Economics in Nairobi, Kenya.[11] Participants began the lab session by familiarizing themselves with the computer keyboard and touchscreen and taking a simple literacy test. They also engaged in training and practice rounds on the DG, SDRT, and combined DG-SDRT. These practice modules were accompanied by comprehension tests to confirm understanding. As preregistered, our main analysis only includes participants who correctly answered at least 5 of the 6 questions on the literacy test, at least 2 of the 3 questions on the DG comprehension test, and at least 3 of the 4 questions on the SDRT practice test. These criteria leave us with 81% of the subject pool (450 of the 558 participants).[12]

After completing all nine of the DG conditions (3 levels of distraction * 3 ethnicity profiles), participants completed an additional DG task designed to test whether the SDRT interrupted their ability to absorb and respond to the information provided in the receiver's profile. Participants were randomly assigned to the standalone DG or SDRT-2 and, following the DG allocation decision, asked a series of multiple choice questions testing their memory of the age, hometown, education level, and name of the receiver whose profile they had just seen.

The lab session concluded with a post-survey including questions about how long the participant has lived in Nairobi, the characteristics of the neighborhood in which they live, their media consumption, their social distance from members of the out-group, and a standard battery of questions measuring motivation to control prejudice.

In order to test additional hypotheses about the conditions under which the SDRT should be most distracting and under which demand effects are likely to be strongest, we randomized the 39 lab sessions to one of four treatment arms that varied whether participants were provided monetary incentives for their accuracy in the SDRT and whether the task instructions on the DG gave participants explicit "permission to discriminate".[13]

[11] We limit the sample to members of just two ethnic groups to maximize study power. We focus on the Kikuyu and Luo to sharpen our focus on politically relevant inter-ethnic differences, as they are the two most historically antagonistic and politically competitive groups in Kenya (Hornsby, 2013).

[12] The literacy screening is important as a means of ensuring the participants were likely able to follow the task instructions as intended. Of the excluded 108, 74 were dropped due to the literacy filter alone, with the other two filters only excluding an additional 34 participants. We report the results of relaxing these standards below.

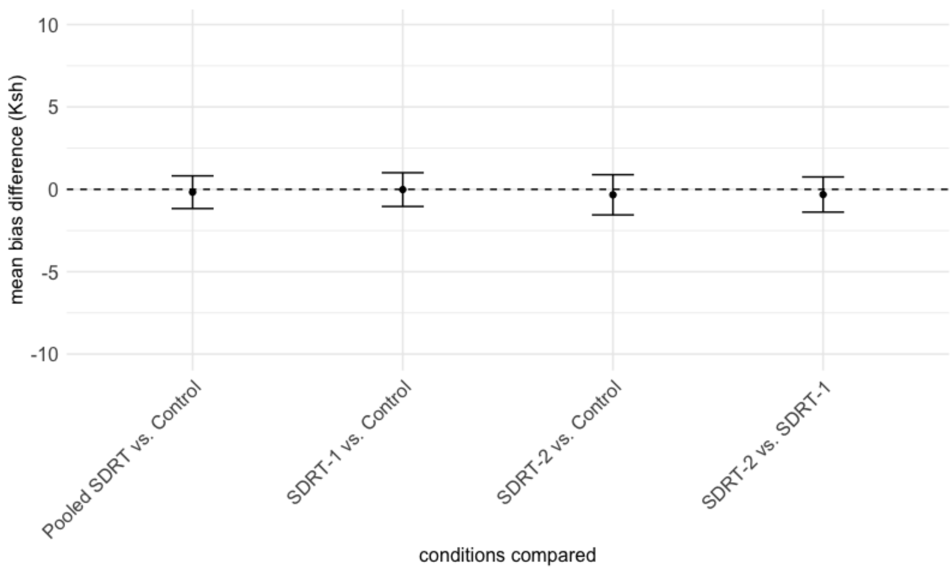
[13] In the incentivized SDRT treatment, participants were told: "In addition to any money you keep from the allocation

Results

Main Findings

Our main, pre-registered results are reported in Figure 3. Contrary to expectations, we find no differences in the sizes of offers to coethnics and non-coethnics—that is, no difference in ethnic bias—across DGs played with and without the distraction task.

Figure 3: Main Results: Differences in Ethnic Bias, by Condition



As shown in Figure 4, DG offers are nearly identical (averaging just below KSh 12, or about a quarter of the KSh 50 endowment) irrespective of cognitive load and irrespective of whether the receiver is a coethnic or non-coethnic (or of indeterminate/neutral ethnic background).[14] These null results are robust to lowering the literacy threshold for inclusion in the sample to just 4 of 6 literacy test questions answered correctly.

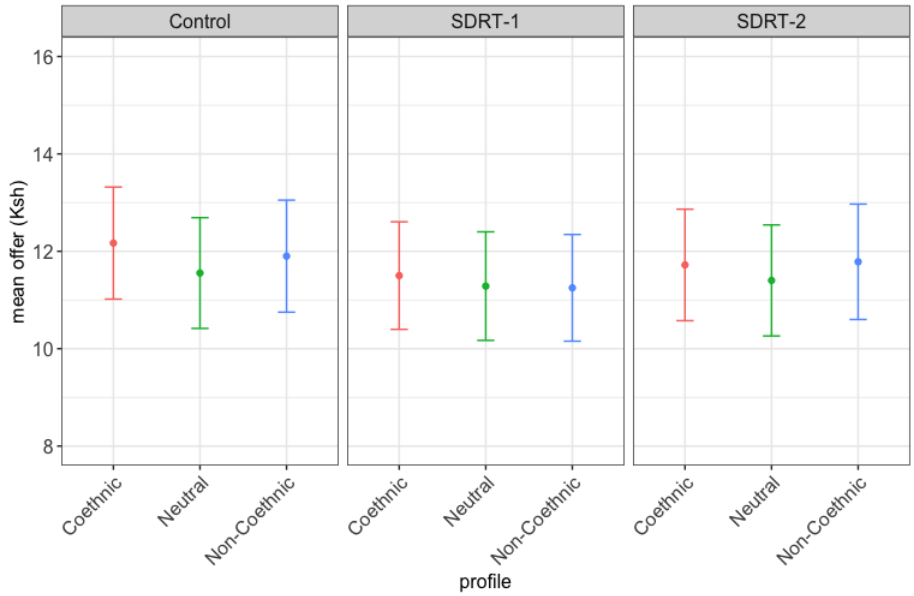
Heterogeneous effects

Beyond these main effects, we designed the study to test whether the impact of the distraction task varied under certain experimental conditions. For example, in keeping with the discussion in de Quidt, Haushofer and Roth (2018), we hypothesized that our findings might differ when participants are ex-

game, you may also earn up to 200 Ksh for the memory task depending on how accurate you are." In the permission to discriminate treatment, participants were told: "Participants from around the world have completed this activity. When participants are given information about the other person, this information often leads them to give more or less money. Participants have many reasons to give more or less money to different people, and that is perfectly okay. How you choose to divide the money is entirely up to you, and whatever decision you make is acceptable."

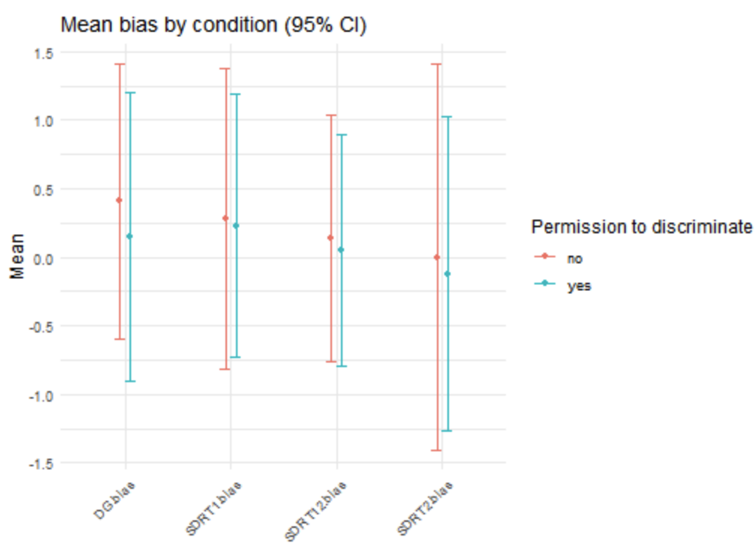
[14] As shown in Appendix A, the distribution of DG allocations is also very similar across receiver profiles and levels of cognitive load.

Figure 4: Main Results: Dictator Game Offers by Condition and Ethnic Profile



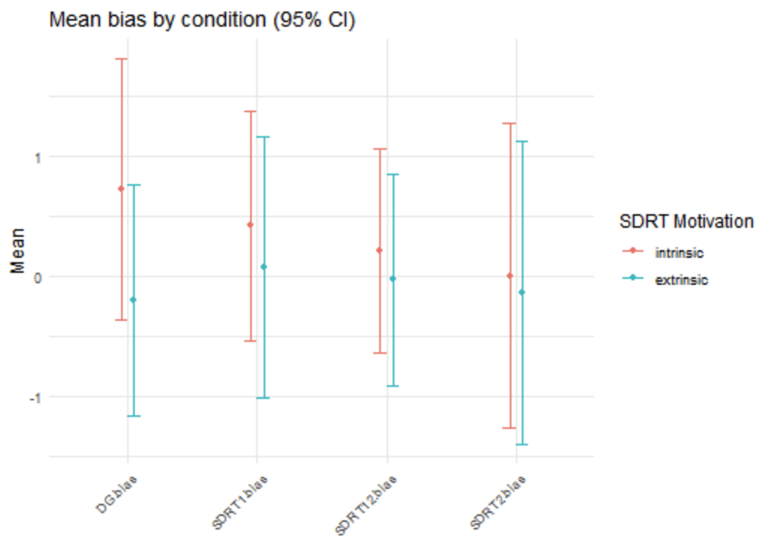
plicitly given permission to discriminate. Such permission might increase ethnic bias in the standalone DG and reduce the bias-revealing effects of the distraction task. As shown in Figure 5, we find no differential effects when participants are explicitly told that discrimination is permitted.

Figure 5: Effect of Giving Participants Permission to Discriminate



Insofar as the hypothesized impact of the SDRT depends on the effort participants invest in remembering the position(s) of the dot(s), we hypothesized that its effects might be greater when participants' were given a financial incentive for their performance on the SDRT. However, as shown in Figure 6, we find no difference in ethnic bias across lab sessions in which participants were incentivized in the SDRT versus where their motivation was purely intrinsic.

Figure 6: Effect of Incentivizing Performance on the SDRT



We also tested whether our findings differed with the characteristics of the participants themselves. Specifically, we investigated whether both ethnic bias and the effects of the distraction task on revealing such bias varied across Kikuyu and Luo participants; with length of residence in Nairobi (which might be correlated with a reduction in ethnic bias and/or socialization to the need to hide such bias (Berge et al., 2020; Kim and Horowitz, 2022; Lyon, 2023)); with the diversity of the participant's neighborhood and reported levels of contact with members of other ethnic groups (Scacco and Warren, 2018); and with the participant's motivation to control prejudice, as measured via a standard scale of questions. As shown in Appendix B, we find no statistically significant differences in either ethnic bias or in the effects of inducing cognitive load for any of these individual-level characteristics.

How Confident Are We That the Distraction Task Worked?

What accounts for these consistent null findings? One possible explanation is that the distraction task failed to knock out the ability to hide one's ethnic biases, either because participants did not exert enough effort in the SDRT or because the task was insufficiently distracting to meaningfully deplete their cognitive control. One version of the latter explanation would hold if the act of hiding one's bias in this setting is so automatic and easy that it is unlikely to be decremented by even a very well calibrated distraction task.

To assess the first possibility, we can draw on data we collected at the end of the lab session about how important participants felt it was for them to remember the position of the dots in the SDRT. Although the vast majority of participants indicate that it had been "very important" (see Appendix C), we do not put too much weight on these findings because participants' responses were almost certainly affected by their desire to say that they had faithfully complied with the task instructions. A more meaningful measure of effort is the share of SDRT trials in which participants were able to correctly indicate whether the probe dot was in the same location as (one of the) cue dot(s). Although this is not a direct indicator of effort (participants could try hard but still get it wrong), it is likely correlated with effort, and also reflects that participants were at least on task, and that the SDRT-1 or SDRT-2 were not inordinately difficult, which might amplify satisficing. If we assume that participants had a 50% chance of guessing correctly on each SDRT trial, we can compare the number of correct responses to chance. When we do, we see strong evidence that participants are doing better than chance in the SDRT-1: the proportion getting all three SDRT-1 responses correct was 2.2 times the chance proportion (exact binomial $p < 1e - 15$). In SDRT-2, far more individuals than expected by chance got *none* of the responses correct (exact binomial $p < 1e - 15$). Our supposition is that this can only occur if participants are in fact responding non-randomly, and we suspect that an error in the way that responses were coded might be responsible for the surprising direction of this difference. We continue to examine this possibility but note for now that non-random behavior in response to the SDRT is statistically a near certainty, suggesting to us that participants likely spent effort on it.

The next question is whether this effort was sufficient to induce cognitive load. To address this issue, we can compare performance on the simple math test that participants took at the beginning of the study, which was done both during the practice rounds of the standalone DG and the SDRT+DG (i.e., both under cognitive load and not).[15] We find small but statistically significant differences in participants' performance on the math test across the two conditions, with worse performance when the test is embedded within the distraction task, indicating that the distraction task does indeed impede cognitive function. Again, we find bigger differences between the standalone DG and SDRT-1 than between the no load condition and SDRT-2 (see Appendix C for details).

A second test of whether the SDRT successfully induced cognitive load comes from the task completed at the end of the lab session to evaluate whether the SDRT undermined participants' abilities to absorb and respond to the information contained in the receiver's profile. As shown in Appendix C, we see little difference across the load and no load conditions in the number of profile elements that were correctly recalled, although we do see significant differences in participants' recollection of one of the four elements (age).

Taken together with the results of the math test, we interpret this as (weak) evidence that the

[15]For evidence that cognitive load reduces performance on math problems, see Deck and Jahedi (2015).

SDRT did in fact induce cognitive load. Combined with the results of our test of effort, we can tentatively conclude that our attempts to experimentally induce cognitive load were successful—and that our failure to find greater ethnic bias in the DGs played concurrently with the distraction task indicate that self-monitoring is not responsible for the failure of DGs to detect cross-group bias.

We are presently undertaking additional analyses to test whether participants who performed better on the SDRT (averaged across all of their trials) gave less in the DG. The logic is that those who perform better on the SDRT are likely trying harder and thus more likely to be under cognitive load when playing the DG—and thus more likely to have their DG allocations affected when completing the task simultaneously with the SDRT.

Given the consistently stronger results in these analyses for SDRT-1 than SDRT-2, it is worth considering whether the SDRT-2 may have been too challenging, causing participants to "give up" and cease allocating attentional resources to the task. We embarked on the project with uncertainty about whether the SDRT-2 task exceeded this difficulty tipping point.[16] While our results here suggest that it might have, we note that our main (null) findings are robust to focusing on the SDRT-1 trials alone, and that we see no statistically significant differences in DG allocations under SDRT-1 and SDRT-2 (see Figure 3).

Interpreting Our Findings

The straightforward interpretation of our findings is that, contrary to expectations, the inability of the DG to detect ethnic bias in a setting where cross-group biases are almost certainly present is *not* because the DG permits self-monitoring and control. This leaves two possible explanations for the puzzle that motivated our investigation.

First, ethnic bias may simply be absent in the setting we study, or too weak to be detected via the DG. This would explain both why we fail to detect cross-group bias in the standalone DG and why there is no difference in measured bias between the standalone DG and the versions of the game where one's ability to mask one's biases is undermined via a distraction task. While the voluminous literature on ethnic conflict in Kenya—a literature that especially highlights the historically fraught relationship between Kikuys and Luos (Gibson and Long, 2009; Wrong, 2009; Hornsby, 2013)—suggests that this explanation is unlikely, we cannot rule it out entirely.

A second potential explanation is that while cross-group biases exist in our study setting, they are not biases of the type that the DG is well-suited to measure. In a separate line of research, Blum, Hazlett and Posner (2021) distinguish between ethnic bias rooted in differential altruism (which the

[16] Indeed, our pilot study included both an SDRT-1 and SDRT-3 task, with most participants performing worse than chance on the latter. For this reason, we replaced the SDRT-3 with the SDRT-2 for the main study. However, the SDRT-2 task, while easier, may still be too difficult for the intended purpose.

DG is specifically designed to capture) and ethnic bias rooted in disliking, fear, and threat perception (which may be better detected via other measurement tools). If the conflictual patterns of inter-group behavior that inform our priors about the level of ethnic bias in Kenya are driven by biases of the latter sort, then this may explain our inability to detect cross-group bias in the DG. This conjecture is supported by the fact that both Berge et al. (2020) and Blum, Hazlett and Posner (2021), who also work in Nairobi and find no evidence of ethnic bias in the DG, *do* find evidence of ethnic bias, respectively, in the Implicit Attitude Test and the Affect Misattribution Procedure—measures designed to capture negative affect toward non-cotethnics. Blum, Hazlett and Posner (2021) also find evidence evidence of ethnic bias in a Weapon Misidentification Task, designed to capture fear and threat perception of out-group members.

The research program of measuring ethnic (and also racial) bias is important to meeting our societal goals. Beyond the competing interpretations for our results, the main lessons of our research program thus far may be the importance of (i) conceptual clarity regarding different notions of racial and ethnic bias that attach to different outcomes, (ii) the compatibility of those different outcomes and different tools for measuring bias across them; and (iii) the interplay of these biases and efforts to control the appearance of bias in our behaviors. Towards this goal, the present study suggests at minimum that manipulating cognitive load as a means to lessen efforts to control bias does not restore a bias of the differential altruism type that the DG is designed to measure.

References

Adida, Claire L., David D. Laitin and Marie-Anne Valfort. 2014. "Muslims in France: Identifying a discriminatory equilibrium." *Journal of Population Economics* 27(4):1039–1086.

Ahmed, Ali M. 2010. "What is in a surname? The role of ethnicity in economic decision making." *Applied Economics* 42(21):2715–2723.

Bardsley, Nicholas. 2008. "Dictator game giving: altruism or artefact?" *Experimental Economics* 11(2):122–133.

Barriga, Alicia, Neil T.N. Ferguson, Nathan Fiala and Martin Alois Leroch. 2023. "Ethnic cooperation and conflict in Kenya." *Journal of Behavioral and Experimental Economics* 106:1–12.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel N. Posner, Bertil Tungodden and Kelly Zhang. 2020. "Ethnically biased? Experimental evidence from Kenya." *Journal of the European Economic Association* 18(1):134–164.

Blum, Ashley, Chad Hazlett and Daniel N Posner. 2021. "Measuring ethnic biases: Can misattribution-based tools from social psychology reveal group biases that economics games cannot?" *Political Analysis* 29(3):385–404.

Chen, Yan and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99(1):431–357.

Cilliers, Jacobus, Oeindrila Dube and Bilal Siddiqi. 2015. "The white-man effect: How foreigner presence affects behavior in experiments." *Journal of Economic Behavior & Organization* 118:397–414.

de Quidt, Jonathan, Johannes Haushofer and Christopher Roth. 2018. "Measuring and bounding experimenter demand." *American Economic Review* 108(11):3266–3302.

Deck, Cary and Salar Jahedi. 2015. "The effect of cognitive load on economic decision making: A survey and new experiments." *European Economic Review* 78:97–119.

Devine, Patricia G. 1989. "Stereotypes and prejudice: Their automatic and controlled components." *Journal of Personality and Social Psychology* 56(1):5.

Eifert, Benn, Edward Miguel and Daniel N. Posner. 2010. "Political competition and ethnic identification in Africa." *American Journal of Political Science* 54(2):495–510.

Fershtman, Chaim and Uri Gneezy. 2001. "Discrimination in a segmented society: An experimental approach." *Quarterly Journal of Economics* 116(1):351–377.

Gibson, Clark C. and James D. Long. 2009. "The presidential and parliamentary elections in Kenya, December 2007." *Electoral Studies* 28(3):497–502.

Glahn, D.C., J. Kim, M.S. Cohen, V.P. Poutanen, S. Therman, S. Bava, T.G.M. Van Erp, M. Manninen, M. Huttunen, J. Lönnqvist et al. 2002. "Maintenance and manipulation in spatial working memory: dissociations in the prefrontal cortex." *Neuroimage* 17(1):201–213.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2009. *Co-ethnicity: Diversity and the dilemmas of collective action*. Russell Sage Foundation.

Haley, Kevin J. and Daniel M.T. Fessler. 2005. "Nobody's watching? Subtle cues affect generosity in an anonymous economic game." *Evolution and Human Behavior* 26(3):245–256.

Hauge, Karen Evelyn, Kjell Arne Brekke, Lars-Olof Johansson, Olof Johansson-Stenman and Henrik Svedster. 2016. "Keeping others in our mind or in our heart? Distribution games under cognitive load." *Experimental Economics* 19:562–576.

Haushofer, Johannes, Sara Lowes, Musau Abednego, David Ndetei, Nathan Nunn, Moritz Poll and Nancy Qian. 2023. "Stress, ethnicity, and prosocial behavior." *Journal of Political Economy, Microeconomics* 1(2):225–269.

Hazlett, Chad, Emilie Campos, Patricia Tan, Holly Truong, Sandra Loo, Charlotte DiStefano and Damla Jeste, Shafali Senturk. 2020. "Principle ERP reduction and analysis: Estimating and using principle ERP waveforms underlying ERPs across tasks, subjects and electrodes." *Under review* .

Hornsby, Charles. 2013. *Kenya: A history since independence*. I.B. Tauris.

Iyengar, Shanto and Sean J. Westwood. 2015. "Fear and loathing across party lines: New evidence on group polarization." *American Journal of Political Science* 59(3):690–707.

James, William. 1890. *The principles of psychology*. Henry Holt.

Jeon, Sangick, Tim Johnson and Amanda Lea Robinson. 2017. "Nationalism and social sanctioning across ethnic lines: Experimental evidence from the Kenya-Tanzania border." *Journal of Experimental Political Science* 4(1):1–20.

Kim, Hue-Sung and Jeremy Horowitz. 2022. "Rejecting ethnic pandering in urban Africa: A survey experiment on voter preferences in Nairobi, Kenya." *Political Research Quarterly* 75(4):1240–1254.

Lenartowicz, Agatha, Holly Truong, Giulia C. Salgari, Robert M. Bilder, James McGough, James T. McCracken and Sandra K. Loo. 2019. "Alpha modulation during working memory encoding predicts neurocognitive impairment in ADHD." *Journal of Child Psychology and Psychiatry* .

Levitt, Steven D. and John A. List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives* 21(2):153–174.

List, John A. 2007. "On the interpretation of giving in dictator games." *Journal of Political Economy* 115(3):482–493.

Lyon, Nicholas. 2023. "Unpacking urban: City size, interethnic trust, and prejudice in Africa." Unpublished paper.

McCauley, John F. 2014. "Pentecostalism as an informal political institution: Experimental evidence from Ghana." *Politics and Religion* 7(4):761–787.

Mironova, Vera and Sam Whitt. 2014. "Ethnicity and altruism after violence: The contact hypothesis in Kosovo." *Journal of Experimental Political Science* 1(2):170–180.

Payne, B. Keith. 2001. "Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon." *Journal of Personality and Social Psychology* 81(2):181.

Payne, B. Keith, Clara Michelle Cheng, Olesya Govorun and Brandon D. Stewart. 2005. "An inkblot for attitudes: affect misattribution as implicit measurement." *Journal of Personality and Social Psychology* 89(3):277.

Peterson, Brenton D. 2016. "Social identification and in-group favoritism: Measuring the strength of group ties." Unpublished paper.

Posner, Daniel N. 2024. "Ethnic bias in urban and rural Africa." Unpublished paper.

Rand, David G., Alexander Peysakhovich, Gordon T. Kraft-Todd, George E. Newman, Owen Wurzbacher, Martin A. Nowak and Joshua D. Greene. 2014. "Social heuristics shape intuitive cooperation." *Nature communications* 5(1):1–12.

Scacco, Alexandra and Shana S. Warren. 2018. "Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria." *American Political Science Review* 112(3):654–677.

Schulz, Jonathan F., Urs Fischbacher, Christian Thoni and Verena Utikal. 2014. "Affect and fairness: Dictator games under cognitive load." *Journal of Economic Psychology* 41:77–87.

Tanaka, Tomomi and Colin F. Camerer. 2016. "Trait perceptions influence economic out-group bias: lab and field evidence from Vietnam." *Experimental Economics* 19(3):513–534.

van der Merwe, Wilhelm Gerhard and Justine Burns. 2008. "What's in a name? Racial identity and altruism in post-Apartheid South Africa." *South African Journal of Economics* 76(2):266–275.

Whitt, Sam and Rick K. Wilson. 2007. "The Dictator Game, fairness and ethnicity in postwar Bosnia." *American Journal of Political Science* 51(3):655–668.

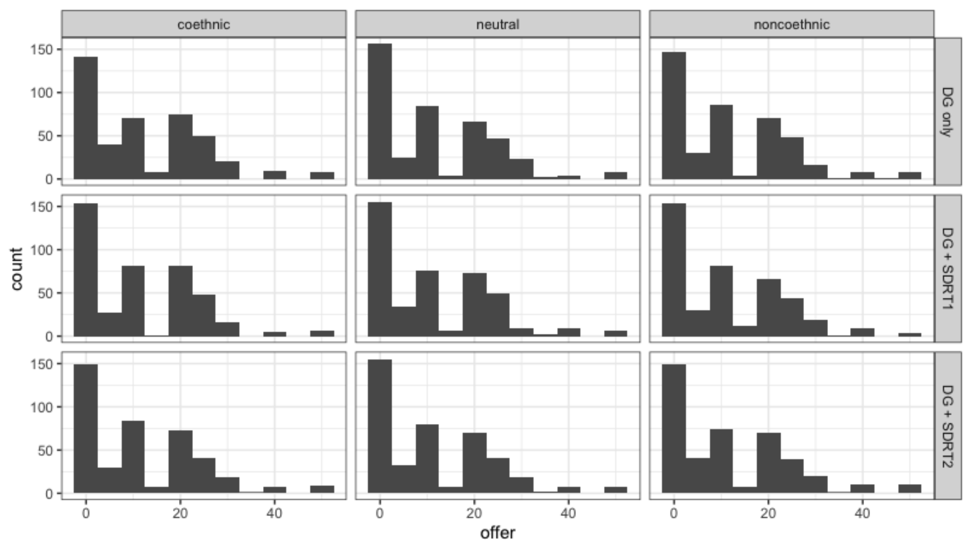
Wrong, Michela. 2009. *It's our turn to eat.* Fourth Estate London.

Zizzo, Daniel John. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics* 13:75–98.

Appendix

A Distribution of DG Allocations

Figure A.1: Distribution of DG Offers, By Profile and Level of Distraction



B Heterogeneous Treatment Effects

C Assessing the Impact of the SDRT

Figure B.1: Differential Effects by Ethnicity

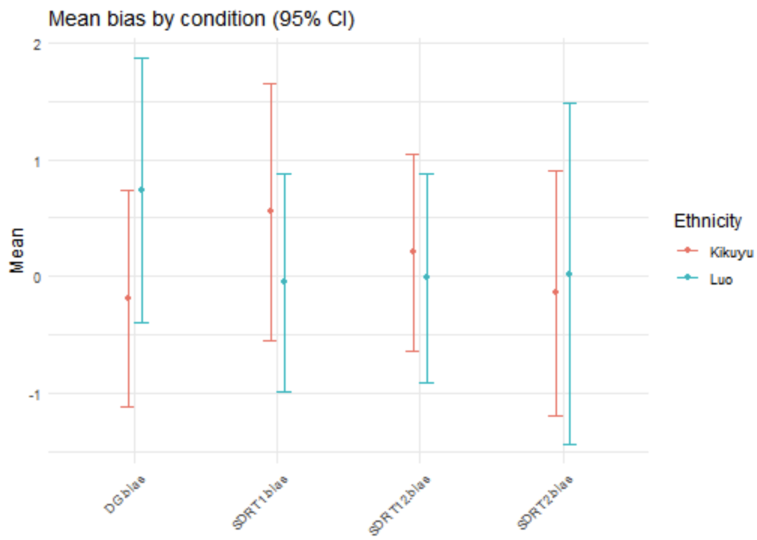


Figure B.2: Differential Effects by Length of Residence in Nairobi

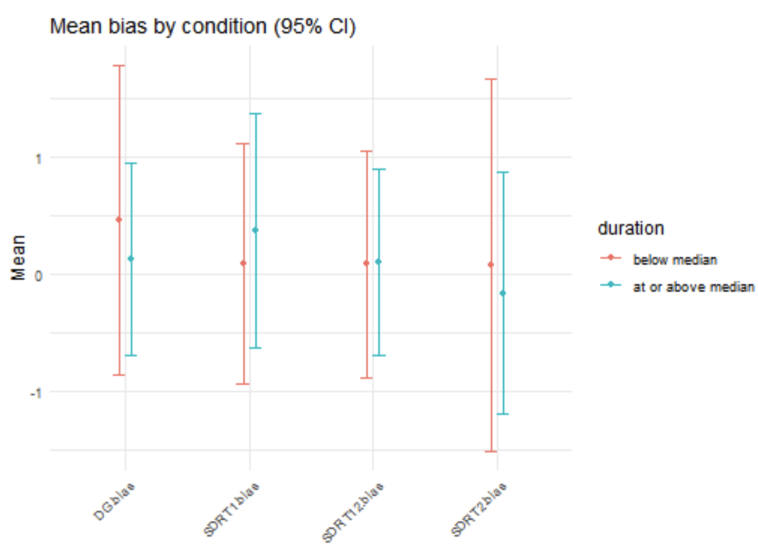


Figure B.3: Differential Effects by the Diversity of the Participant's Neighborhood

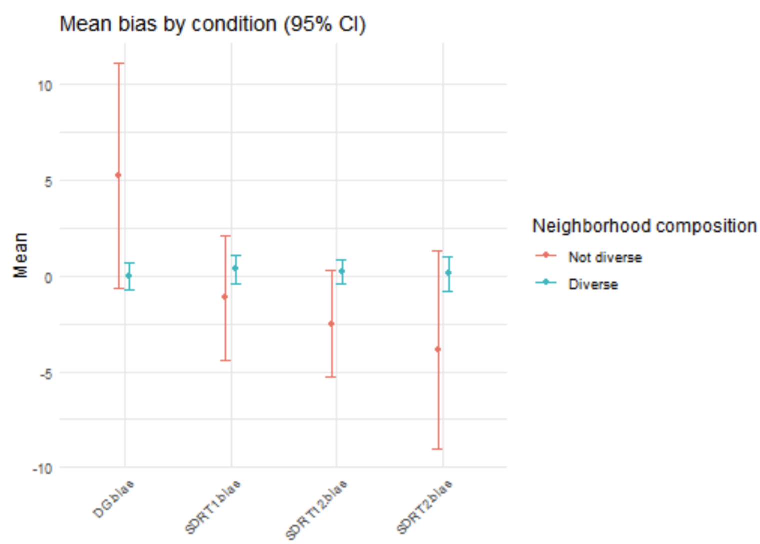


Figure B.4: Differential Effects by Reported Contact with Non-coethnics

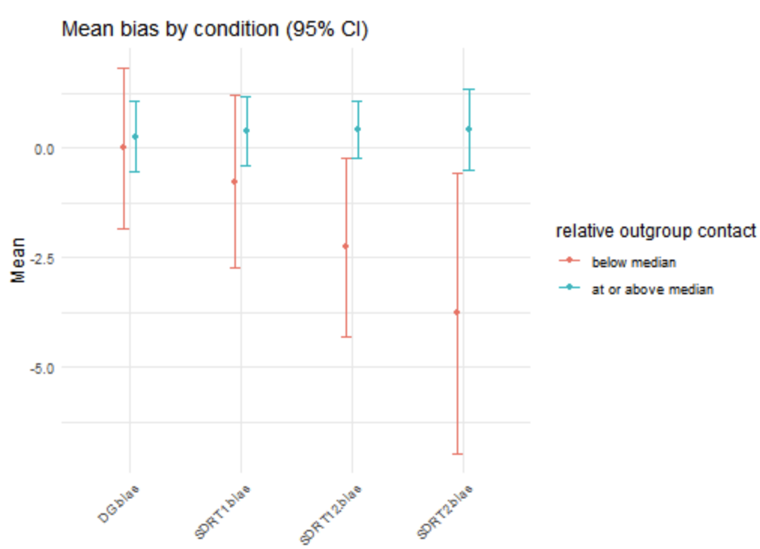


Figure B.5: Differential Effects by Motivation to Control Prejudice

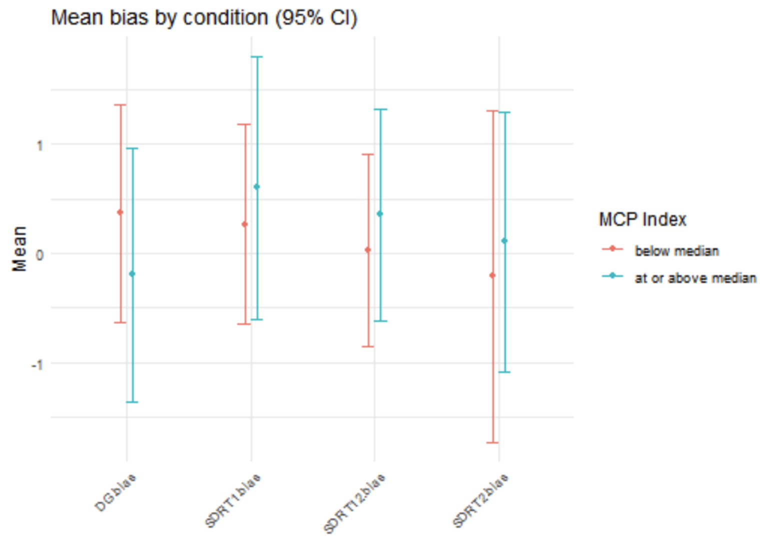


Figure C.1: Reported Effort in the SDRT

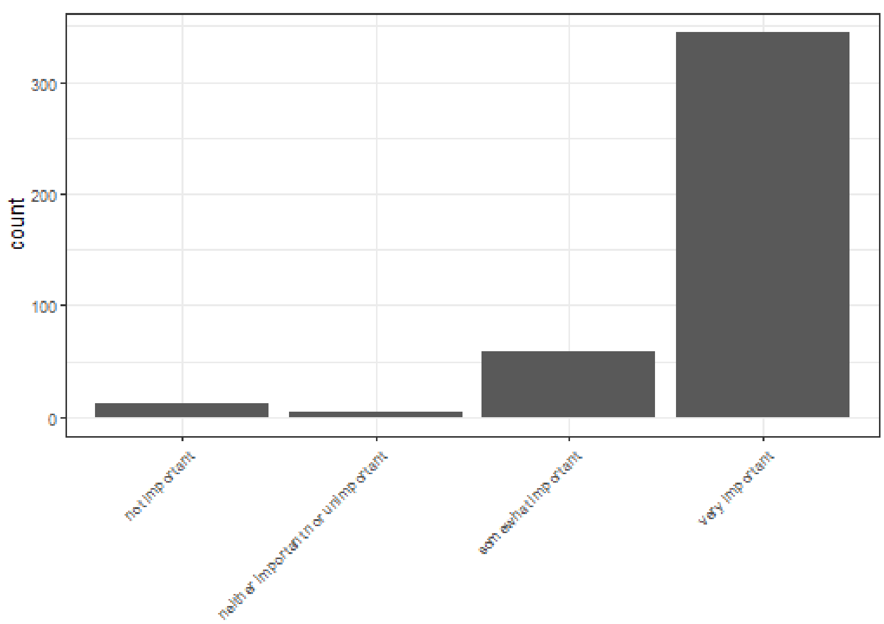


Figure C.2: Effect of the SDRT on Profile Processing