



# OPEN Explicit discrimination and ingroup favoritism, but no implicit biases in hypothetical triage decisions during COVID-19

Nico Gradwohl<sup>1,2,3✉</sup>, Hansjörg Neth<sup>1,2</sup>, Helge Giese<sup>1,2,4</sup> & Wolfgang Gaissmaier<sup>1,2</sup>

Disturbingly realistic triage scenarios during the COVID-19 pandemic provide an opportunity for studying discrimination in moral reasoning. Biases and favoritism do not need to be explicit and overt, but can remain implicit and covert. In addition to assessing laypeople's propensity for engaging in overt discrimination, the present study examines whether they reveal implicit biases through seemingly fair random allocations. We present a cross-sectional online study comprising 6 timepoints and a total of 2296 participants. Each individual evaluated 19 hypothetical scenarios that provide an allocation dilemma between two patients who are in need of ventilation and differ only in one focal feature. Participants could either allocate the last ventilator to a patient, or opt for random allocation to express impartiality. Overall, participants exhibited clear biases for the patient who was expected to be favored based on health factors, previous ethical or caretaking behaviors, and in-group favoritism. If one patient had been pre-allocated care, a higher probability of keeping the ventilator for the favored patient indicates persistent favoritism. Surprisingly, the absence of an asymmetry in random allocations indicates the absence of covert discrimination. Our results demonstrate that laypeople's hypothetical triage decisions discriminate overtly and show explicit biases.

A shortage of intensive care capacities during the COVID-19 pandemic forced doctors to make disturbing decisions in which they had to prioritize which patient to save<sup>1,2</sup>. Laypeople's decisions in such triage scenarios<sup>3</sup> provide an opportunity for studying moral reasoning in less artificial contexts than traditional sacrificial dilemmas<sup>4–7</sup>, such as the trolley dilemma<sup>8,9</sup>. Recent ethical guidelines for medical situations of extreme scarcity are mostly based on utilitarian considerations. To save the largest number of lives, they advocate saving those patients with the best prognosis<sup>2</sup>. In the absence of relevant distinctions, ethical guidelines recommend a random allocation of resources<sup>1,10</sup> (but see<sup>11</sup>). Laypeople largely endorse such guidelines, but have been found to deviate from random allocations by incorporating and overtly discriminating based on a variety of patient features that should be ignored from a normative stance<sup>12–19</sup>. Although people disagree with discriminating policies<sup>12</sup>, judgments of concrete cases and abstract principles may diverge (e.g.<sup>20</sup>), explaining widespread evidence for discrimination in other work<sup>13,14</sup>. Beyond that, focusing on overt measures of discrimination may still underestimate the extent of biases and in-group favoritism due to implicit biases (i.e., biases that are not revealed when people are asked explicitly), social desirability, and under-reporting effects<sup>21,22</sup>. Thus, we move beyond assessing laypeople's propensity for engaging in overt discrimination and study whether they reveal further biases by allowing them to use seemingly fair random allocations to discriminate covertly against less favored patients.

It is well-established that people's judgments and decisions rarely adhere strictly to norms of moral reasoning<sup>20,23–25</sup>. People deviate from utilitarian norms of impartiality (e.g.<sup>20</sup>) and egalitarian norms of equality<sup>24</sup>, and aim to avoid harming others, even if harming others would save a larger number of lives overall<sup>23,25</sup>. Laypeople also show a variety of pervasive biases, many of which align with their perceptions of how scarce medical resources should be allocated during crises (see<sup>14</sup>, for a review), and lead to systematic discrimination. Biases for health and longevity generally favor younger and healthier individuals, as shown in studies of QALY maximization (see<sup>26</sup>, for a review), survey experiments on the allocation of scarce medical resources<sup>24,27</sup> and hypothetical triage scenarios during the COVID-19 pandemic<sup>13–16,18,19,28</sup>. Individuals are especially disfavored if they

<sup>1</sup>Department of Psychology, University of Konstanz, Konstanz, Germany. <sup>2</sup>Cluster of Excellence "Centre for the Advanced Study of Collective Behaviour", University of Konstanz, Konstanz, Germany. <sup>3</sup>Department for the Ecology of Animal Societies, Max Planck Institute of Animal Behavior, Konstanz, Germany. <sup>4</sup>Charité – Universitätsmedizin Berlin, Berlin, Germany. ✉email: nico.gradwohl@uni-konstanz.de

are perceived to be responsible for their ill-health, as in the case of obese individuals<sup>29</sup> or with other medical conditions<sup>13,18</sup> (but see<sup>30</sup>). People also tend to favor individuals who exhibit behaviors that are being perceived as ethical, are being described as cooperative and as contributing to society<sup>31</sup>. Similarly, patients who refrain from delinquent behaviors<sup>13,16</sup>, got vaccinated<sup>32</sup>, or abstained from risky (health) behavior<sup>18</sup> are more likely to be favored. Relatedly, decision makers in triage decisions are willing to consider family-ties and caretaking roles, favoring patients with dependents<sup>13,19</sup>.

Individuals generally tend to cooperate more willingly with related others<sup>33</sup> and with others that are perceived to belong to one's in-group in economic games, such as the Prisoner's dilemma or Dictator games<sup>34–38</sup>. These biases and in-group favoritism permeate into moral judgments and decisions and affect who would be saved in sacrificial dilemmas<sup>24,39–44</sup>. Such dilemmas include trolley-type dilemmas with self-driving cars<sup>45</sup>, but also triage situations involving the allocation of scarce healthcare resources<sup>46</sup>, where people also base their preferences on patients' ethnic background<sup>19</sup>, citizenship<sup>13,47</sup>, whether patients recently migrated<sup>48</sup>, political affiliations (but see<sup>49</sup>), and shared nationality, especially when decision makers were more politically conservative<sup>16</sup>.

Manifestations of biases and favoritism need not be explicit and overt, but can remain implicit and operate in a covert fashion<sup>50–52</sup>, affecting both the offers in trust games<sup>53</sup> and sacrificial dilemma decisions<sup>54</sup>. Related to modern forms of racism (e.g.,<sup>55</sup>), people supporting egalitarian norms more readily exhibit favoritism in situations that allow them to justify it, for instance, by punishing outgroup members more readily by rejecting unfair offers<sup>56–58</sup>. Moreover, people strategically use seemingly fair equality norms—such as randomness—in self-serving ways. They use random allocations to avoid responsibility for their outcomes<sup>59</sup>, and prefer the use of random devices to create plausible deniability<sup>60</sup>. Similarly, people use the alleged outcome of a coin flip to justify unequal allocations<sup>61</sup> and judge equal allocations as less fair if an individual from their group was previously favored by an allocation decision<sup>62</sup>. Thus, people could use seemingly fair principles of random allocation to avoid responsibility, but also to justify biased decisions.

As a facet of avoiding responsibility, a bias for inaction that favors omission over commission<sup>63,64</sup> has been linked to the avoidance of harm in sacrificial dilemmas<sup>8,9,23</sup>. In triage decisions, people are generally reluctant to withdraw care from patients who already receive treatment<sup>12,15,19,32</sup>. Effectively, this tendency manifests itself in a first-come, first-served principle, that is discouraged by ethical guidelines<sup>1</sup>. In contrast to this reluctance to withdraw, laypeople are more willing to harm a patient by withdrawing care from a patient with a worse medical prognosis<sup>19</sup>. Thus, deviations from norms of moral reasoning may be based on more general mechanisms of judgment and decision making, including in-group favoritism and omission bias, that may also affect the allocation of scarce medical resources.

Prior research on triage decisions primarily focused on explicit biases and overt discrimination. This focus may be prone to underestimate the true extent of biases and in-group favoritism<sup>21,22</sup>. Beyond replicating and extending the evidence for overt discrimination, we also examine covert discrimination tendencies. A diagnostic case for comparing covert and overt discrimination decisions is provided by situations in which a pre-allocated medical resource can be re-allocated. When withdrawing care causes active harm, strategically advocating random re-allocation provides an opportunity for covert discrimination that may appear to adhere to ethical guidelines. Rather than explicitly showing a bias by overtly reallocating a ventilator to a favored patient, a higher willingness to opt for random allocation when a ventilator has been pre-allocated to a disfavored patient than when it has been pre-allocated to a favored patient may superficially satisfy equality norms or ethical guidelines<sup>1,10</sup>. However, asymmetric use of random re-allocation would nonetheless violate impartiality by selectively increasing the survival chances of a favored patient, relative to keeping the ventilator pre-allocated to the unfavored patient. Thus, an asymmetric use of random re-allocation would provide an opportunity for covert discrimination, as it allows individuals to justify their choice and thus avoid blame by choosing a seemingly fair option<sup>59</sup>. Previous studies relying on forced choice experiments, conjoint analysis (estimating the relevance of multiple features from the responses to forced binary choices<sup>16,18</sup>), and ratings of moral relevance<sup>13,48</sup>, either did not allow individuals to explicitly abstain from a decision, or did not provide them with an opportunity for covert discrimination by avoiding direct harm.

The present study investigates whether laypersons advocate an overt allocation of scarce medical resources based on biases and in-group favoritism or covertly express their preferences through seemingly fair random re-allocation decisions. Participants decided for 19 patient pairs how a doctor should allocate the last remaining ventilator to one of two hypothetical patients during a viral pandemic. Since both patients were described to be in critical condition, saving one patient implied sacrificing the other patient. Each pair of patients differed only in one focal feature. As it was emphasized that patients did not differ in any other features (including medical prognosis), an impartial random ventilator allocation was the recommended normative response<sup>1,10</sup>. In contrast to most prior work on triage decisions, participants could always opt for random ventilator allocation (see<sup>19</sup> for a similar approach). Using this setup enables us to disentangle two aspects of discrimination: first, comparing pairs of patients that differ in one distinctive feature allows for the assessment of interpersonal agreement on the bias towards favored vs. dis-favored patient characteristics. Second, stronger deviations from random allocations indicate more pervasive biases, whereas the extent of random allocations provides a measure of impartiality. To distinguish explicit biases in allocation decisions (by denying care to a disfavored patient) from implicit biases reflected in random re-allocation, we moved beyond the basic withholding condition (i.e., both patients arriving simultaneously), and included a withdrawal condition (i.e., one patient arriving before the other) at two time-points (3 and 4, out of a total of 6). In the withdrawal condition, the last ventilator had been pre-allocated to the patient who had arrived earlier, but this default could be revised by re-allocating it to the other patient or by opting for random re-allocation. This extends a similar study by Wilkinson et al.<sup>19</sup>, who also allowed for random allocation and used a withdrawal condition for patients with different survival probabilities. By varying the withdrawal condition for features that are orthogonal to patients' survival probability, our study is the first to address covert discrimination in this setting. This novel approach to study covert discrimination reveals

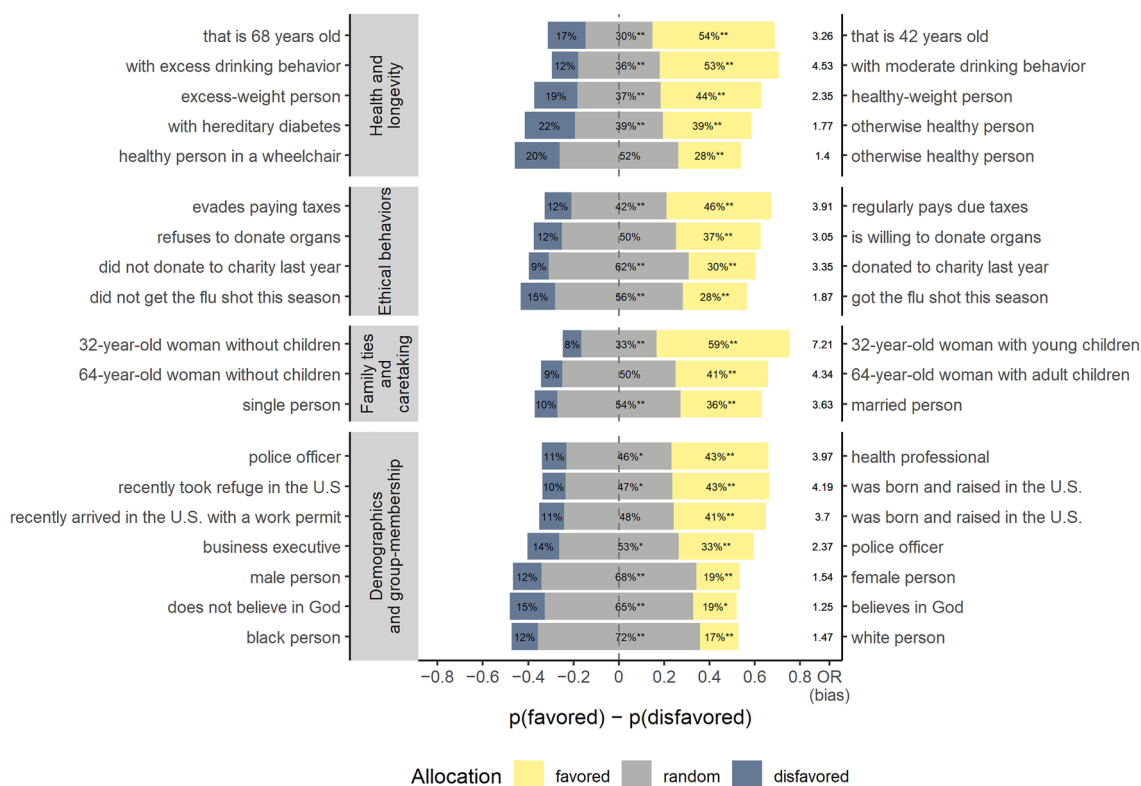
systematic and strong explicit biases towards the patients with features that were expected to be favored. This overt discrimination largely persists under withdrawal conditions, but we find no evidence for implicit biases revealed through the (mis-)use of randomness that could additionally harm disfavored individuals.

## Results and discussion

### Overt discrimination based on health and longevity, ethical behavior, family and caretaking, and demographics

Across all features in the withholding condition (when both patients arrived simultaneously) we find evidence for feature-based discrimination on all 6 timepoints of data collection. Among the non-random allocations, we see strong agreement on a bias towards the patient with features that were expected to be favored [ $p(\text{favored patient}|\text{random allocation}) = 73.95\%$ ]. An average of 49.51% of nonrandom allocations across all features demonstrate a clear lack of an overall agreement on honoring impartiality. Thus, participants mostly agreed when expressing favoritism in the expected direction, but only a minority made consistently impartial allocation decisions (4.20% of participants always used random allocations, 7.20% of participants never did). At the same time, the extent of overt discrimination varied considerably between our 4 categories of patient features: health and longevity, ethical behavior, family and caretaking roles, and demographics. Figure 1 illustrates both the lack of impartiality (by random allocations, shown as grey bars, being substantially below 100%) and the biases towards favored patients (by more overt choices of favored features, shown in yellow, than of disfavored features, shown in blue) (results are largely robust to differences between time-points, see “Materials and methods” and Supplementary Sect. 2.1.3).

Figure 1 reveals interesting similarities and differences regarding the category and specific content of discriminating patient features (feature comparisons compare estimates of bias and impartiality from logistic mixed models and are controlled for all 6 timepoints of data collection, see Supplementary Table 4 for the models and Supplementary Sect. 2.1.3 for an exploration of an interaction with sets of time-points of data collection). Despite our emphasis that each pair of patients only differed in one central characteristic (i.e., not in their survival probability), participants showed particularly pervasive biases on most *health and longevity* features. Among features



**Figure 1.** Discrimination reflected in proportions of random allocations vs. allocations to the favored and disfavored patient by patient feature. The right label corresponds to the patient who we predicted to be favored, and the left label to the patient who we predicted to be disfavored. All four feature clusters indicate that allocations varied between patient features. Asterisks on random allocations indicate that random allocations statistically deviated from 50%, reflecting a lack of impartiality, and asterisks on favored allocations indicate a higher likelihood for favored than for disfavored allocations among non-random allocations, reflecting systematic bias towards one patient (binomial tests against 50%, first asterisk uncorrected, second asterisk Bonferroni corrected for 19 tests). The offset of bars from the center indicate the degree of bias as the absolute difference between favored and disfavored allocations (in percent) and odds ratios (to the right of the bars) provide the effect size of each bias.

for which patient control is likely perceived to be low, there was a comparable amount of bias for a patient with hereditary diabetes and a patient in a wheelchair ( $OR = 0.96$ ,  $p = 0.732$ ), but impartiality towards the patient with diabetes was lower ( $OR = 0.45$ ,  $p < 0.001$ ). Comparing age to hereditary diabetes, in turn, shows both an increased bias ( $OR = 1.94$ ,  $p < 0.001$ ) and lower impartiality ( $OR = 0.55$ ,  $p < 0.001$ ) for age. This replicates earlier findings that patient age is used to discriminate in favor of a younger person<sup>13–16,18,19,28</sup> (but see<sup>30</sup>; supplementary analyses also show no systematic evidence for ageism among younger participants, see Supplementary Sect. 2.3). When considering presumably more controllable features, excess drinking and excess weight elicited stronger biases than hereditary diabetes ( $OR = 3.07$ ,  $p < 0.001$  and  $OR = 1.46$ ,  $p = 0.002$  respectively), while bias against the excess-drinking patient was larger than against the excess-weight patient ( $OR = 2.11$ ,  $p < 0.001$ ). When contrasting the feature age with less controllable alcohol consumption, relatively lower bias for age ( $OR = 0.63$ ,  $p < 0.001$ ) could reflect the co-existence of general ageism with either prioritarian considerations for saving individuals in higher need<sup>12</sup> or merit-based considerations in favor of the elderly<sup>31</sup>, that are absent for alcohol consumption. Conversely, decreased impartiality for age ( $OR = 0.66$ ,  $p < 0.001$ ) reflects the higher willingness to use age as a feature for these different reasons. Overall, these findings are in line with previous research findings that the elderly and individuals possibly responsible for their ill health (e.g., smokers, unvaccinated individuals, and those engaging in risky behaviors) were less likely to be saved<sup>13–16,18</sup>. Thus, perceptions of life-quality and individual control jointly play a role in resource allocation decisions.

With respect to *ethical behaviors*, we find consistent and clear biases towards patients that are being described as behaving more ethically (see Fig. 1). When patients differed in their charitable behavior and flu vaccination status, systematic choices of random allocations indicate that a majority of people honored impartiality in those cases (61.7% and 56.2%, respectively). However, this consensus on impartiality declined when one patient was described as an organ donor (50.4%) and further deteriorated when a patient was described to evade paying taxes (42.0%). Relative to the non-cooperative and delinquent case of tax evasion, the evidence for explicit biases based on other ethical features (e.g., vaccination and organ donation status) were weaker, presumably because they involve personal decisions without violating existing laws. Nevertheless, in-group effects (see Supplementary Sect. 2.3) suggest that those participants who reported the cooperative behavior themselves discriminated more in favor of the cooperative patient, possibly indicating instances of third-party punishment (e.g., like in the case of vaccination<sup>32</sup>).

Patient features implying *family-ties and caretaking roles* elicit substantive discrimination towards individuals who are embedded in family structures and occupy caretaking roles. The extent of biases increases and impartiality decreases from past caretaking to present caretaking. Regarding bias, we find no differences between married patients and those with adult children ( $OR = 1.03$ ,  $p = 0.857$ ), but increased biases towards patients with young children when compared to those with adult children ( $OR = 1.64$ ,  $p = 0.001$ ). However, participants became less impartial towards married patients than towards those with adult children (past caretaking,  $OR = 0.76$ ,  $p = 0.003$ ), and even less impartial towards patients with young rather than adult children (present caretaking,  $OR = 0.36$ ,  $p < 0.001$ ). Strong favoritism for saving a mother with young children corroborates previous findings that patients with dependents are more likely to be saved<sup>13,65,66</sup>. The decrease in impartiality from being married over past caretaking may indicate that contributing potential children in the future raises a patient's perceived deservingness (e.g.,<sup>13</sup>), but not to the extent of already having contributed children in the past.

Finally, considering patients' demographics and group membership, weak biases based on patients' sex, ethnicity, and religiosity, and high impartiality on those features suggest that there were few signs of overt sexism or racism (but the observed biases can partly be related to in-group favoritism, see Supplementary Sect. 2.3). However, pronounced biases towards saving police officers and health professionals indicate favoritism for higher reputation jobs, thereby favoring health-professionals over police officers and business executives. Comparing health professionals to police officers shows a higher bias ( $OR = 1.86$ ,  $p < 0.001$ ) and lower impartiality ( $OR = 0.69$ ,  $p < 0.001$ ) for the former. This may reflect the instrumental value of healthcare workers during the pandemic, as reported in previous work<sup>14,19,65</sup> and reflected in ethical guidelines<sup>1,10</sup>. With respect to residency status, there was substantive bias for patients born and raised in the U.S. over patients described as refugees or work permit holders, although both features lacked consensus on impartiality. Similarly, previous work reported favoritism towards citizens<sup>13</sup>, non-immigrants<sup>47,48</sup>, and patients of identical nationality<sup>16,19</sup>. If such expressions of favoritism were based on past rather than future contributions to society, we would expect work permit holders (who possibly could contribute soon) to be treated more favorably than refugees (who may not as easily contribute). However, neither was the bias for the individual born in the U.S. larger when compared to refugees rather than work permit holders ( $OR = 1.15$ ,  $p = 0.348$ ), nor was the corresponding impartiality decreased ( $OR = 0.93$ ,  $p = 0.449$ ). Participants' reasoning could involve that those born and raised in the U.S. (or their parents) are already contributing to society (e.g., by paying taxes, see the role of tax evasion above). Alternatively, participants' reasoning may reflect in-group favoritism or even racism that is openly shown, as it may appear to be justified based on perceived deservingness or past reciprocity<sup>18,31</sup>. In line with in-group favoritism, we find that impartiality was increased among more liberal participants, indicating that fewer of them discriminated (see Supplementary Sect. 2.3).

Although our pattern of results is largely unchanged upon in- or exclusion of the latter timepoints, minor differences (see Supplementary Sect. 2.3.1) may hint at changes in the participant pool or attitudes towards triage decisions over the course of the pandemic. That a trend towards lower bias holds across patient features could imply that all participants became less willing to rely on random allocations, which was largely driven by increased allocations to the unfavored patient (thus reducing overall bias). Moreover, some of the reported effects seem to be partly driven by participants' group membership and thus reflect in-group favoritism (see Supplementary Sect. 2.3). Such in-group effects affected only the magnitude of bias or impartiality measures, but never reversed the direction of the bias among participants sharing the features of the unfavored patient (with the exception of non-religious individuals, who exhibited a bias against religious patients). Thus, even when

incorporating in-group favoritism, participants' discrimination patterns are driven by overarching preferences for the favored patient.

### Overt but no covert discrimination: No sign of asymmetric re-allocation through randomness

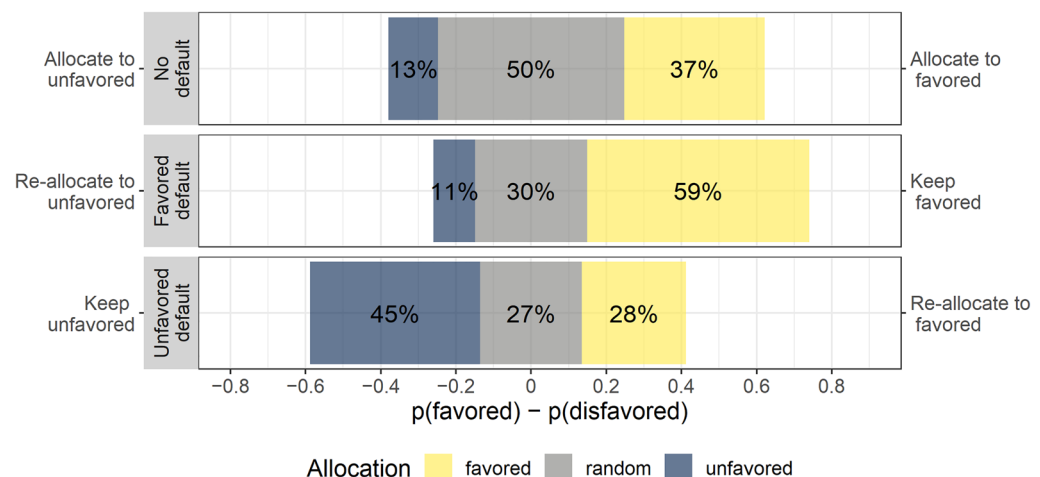
Given clear explicit biases on all of our patient features and a wide-spread willingness to deviate from impartial resource allocation, we now turn to potential implicit biases, as operationalized by the withdrawal conditions (in which the last ventilator had been pre-allocated by default to either a favored or unfavored patient who arrived earlier).

Figure 2 illustrates the willingness for allocating the ventilator in the withholding condition versus keeping it allocated or re-allocating it in the two possible withdrawal conditions. Overall, the allocation probability differed between conditions (according to a logistic mixed model with random effects for participants and patient features,  $\chi^2(2) = 167.858$ ,  $p < 0.001$ ). The default patient was more likely to keep the ventilator than it was allocated to the favored patient in the withholding condition for both, favored or unfavored defaults ( $OR = 4.77$ ,  $p < 0.001$  and  $OR = 1.63$ ,  $p < 0.001$ , respectively). This shows a strong default effect. A higher probability of keeping the ventilator for the favored default than for the unfavored default ( $OR = 2.93$ ,  $p < 0.001$ ), indicates that the favoritism observed in situations of withholding generalized to situations of withdrawal. This extends the previously documented effects regarding resource withdrawal based on differential survival probability<sup>19</sup> and whether a patient was unvaccinated against COVID-19<sup>32</sup>.

Selectively opting for more random re-allocations for less favored patients provided participants with an opportunity for covert discrimination. Although we find an overall effect indicating a difference in random re-allocations between the withdrawal and withholding conditions [ $\chi^2(2) = 399.920$ ,  $p < 0.001$ ], participants' choices for seemingly impartial random re-allocations were not asymmetrically more likely for disfavored defaults ( $OR = 1.17$ ,  $p = 0.562$ , with more random re-allocations towards the favored default). The overall effect merely reflects fewer random re-allocations relative to those in the withholding condition for both favored ( $OR = 0.15$ ,  $p < 0.001$ ) and unfavored defaults ( $OR = 0.13$ ,  $p < 0.001$ ). Consequently, we find no evidence that people systematically used egalitarian random allocations to rationalize their preference against an unfavored default by asymmetrically increasing seemingly fair random allocations, as reported for other settings<sup>59,61</sup>. Instead of engaging in covert discrimination, many participants appeared to use a first-come first-served rule, but still endorsed overt discrimination against the unfavored patient (with the collapsed pattern being robust for all patient features, and for different times of data collection, see Supplementary Sects. 2.2.1 and 2.2.3).

### General discussion

Across a variety of different patient features, we investigated how laypeople decided to withhold and withdraw care in the context of triage decisions. Replicating and extending previous work, participants' withholding decisions showed overt discrimination. They predictably favored more healthy patients and patients with higher status jobs, like health professionals and police officers, but also prioritized patients who are socially embedded and in care-taking roles, as well as patients showing cooperative behaviors. Surprisingly, overt discrimination persisted in withdrawal situations in which a life-saving ventilator could be re-allocated. Rather than using random allocations to covertly express implicit biases, participants consistently made overt re-allocation decisions. Overall, pervasive explicit biases based on patient features profoundly shaped laypersons' triage decisions, even in situations where the decision concerned the withdrawal of care.



**Figure 2.** Probability of receiving the ventilator by condition across patient features. Probabilities to allocate the ventilator to either patients in withholding scenarios (top) vs. retaining or re-allocating it in a withdrawal scenario in which the ventilator has been pre-allocated to a favored default (middle) or to an unfavored default (bottom).

The pervasiveness of discrimination adds to a body of findings that laypersons do not strictly adhere to ethical guidelines<sup>13–16,19,32,65,66</sup> that focus on patient prognosis<sup>2,11</sup> and otherwise recommend random resource allocation<sup>1,10</sup>. Analogous to judgments about whether social welfare is deserved<sup>31</sup>, laypeople consider patients as deserving care to different degrees, based on their perceived level of control over the situation, but also based on societal contributions. Our results on excess weight and alcohol consumption add to mixed findings from conjoint experiments which either find<sup>18</sup> or do not find<sup>30</sup> evidence for these (seemingly) controllable features, but agree on the importance of different professions. Likewise, our findings on patients who recently migrated, but also on marriage, adult children, and young dependents could also reflect a judgment of deservingness due to past or future contributions to society. Interestingly, participants here incorporated features that they had indicated to be irrelevant when asked to judge their moral relevance<sup>13</sup>. They also discriminated based on patients' social usefulness, although they largely rejected discriminating triage policies<sup>12</sup>. This likely implies potential differences between judgements of abstract policies and concrete hypothetical decisions<sup>20</sup>, which could partially consist in opportunities to rationalize discrimination based on features of the specific situation. Additional analyses also show that allocation decisions also reflect the pervasive phenomenon of in-group favoritism<sup>34–38</sup>. The relevance of in-group favoritism for triage decisions has so far only been shown for nationality<sup>16</sup> and cooperative behaviors, such as getting vaccinated<sup>32</sup>. Strikingly, discrimination persisted in situations in which individuals had to decide about actively harming a patient by re-allocating a critical resource. However, people did not seize the opportunity to rationalize their favoritism through seemingly fair random allocations in withdrawal conditions. This is surprising for two reasons: first, people reliably seem reluctant to withdraw care in triage scenarios<sup>12,15,19,32</sup>, avoid directly harming others in sacrificial dilemmas<sup>8,9,23,25</sup>, and exhibit a general preference for inaction over action<sup>63,64</sup>. Second, the observed absence of covert discrimination is at odds with previous reports of the motivated use of randomness<sup>59–61</sup> and people's use of different norms of fairness on outgroup individuals<sup>56,57</sup>.

We cannot fully exclude that experimenter demand characteristics due to salient features or a general aversion against randomness<sup>67</sup> may have led to an overestimation of the degree of discrimination, whereas for some features random allocations may have been more likely due to social desirability (e.g., ethnicity and sex). However, demand effects have been found to be lower in online experiments than in laboratory settings<sup>68</sup>. A general avoidance of random allocations also conflicts with the fact that random allocations were typically the modal response in situations of withholding. Additionally, our discrimination patterns persisted in the withdrawal condition where there was no individual accountability for covert discrimination, which would reduce the potential impact of social desirability. Thus, even though the estimated probabilities of discrimination may be overestimated due to demand characteristics or randomness aversion, the rank order of features should be unaffected and reveal actual preferences. Participants' proneness to openly show their favoritism, even in situations of withdrawal extends the mixed evidence that people are more willing to reallocate a resource from a patient unvaccinated against COVID-19<sup>32</sup> or with a lower survival probability<sup>15,19</sup>. In sum, discrimination appears to be pervasive and extends beyond features that can be aligned with ethical guidelines, like health-related outcomes and a person's instrumental value in conditions of scarcity.

Overall, investigations of hypothetical triage scenarios offer a valuable window into human moral reasoning, which may be subject to a variety of moral motives<sup>43</sup>. The willingness to engage in trade-offs between patient features and situational aspects is in line with broadly utilitarian cost–benefit calculations<sup>69</sup>, but violates norms of impartiality<sup>20</sup> and deontologic rules against harm<sup>20,25,70</sup>. Future work should aim to improve our understanding of the conditions under which people rely on overt and covert forms of discrimination. Although here we were interested in comparing isolated features to get a clean estimate of which features are used, investigating patients described with feature combinations could reveal additional insights, such as trade-offs or interactions between features. This would also be beneficial for external validity, because typically patients will vary in several features. Our results on single features add valuable insights to laypeople's perspectives on practical debates about triage scenarios<sup>14</sup>. Rather than directly translating into guidelines, participants' willingness to overtly express biases and in-group favoritism provides a cautionary note: Results reveal that there are substantial discrepancies between ethical guidelines and people's moral intuitions, resulting in challenging disagreements over particular decisions. At the same time, the fact that people's biases are openly stated rather than covertly enacted enables policy makers to anticipate such discrepancies, allowing—but also requiring—they to better explain expert recommendations to the general public.

## Materials and methods

### Participants

In total, we recruited 2463 participants on Amazon Mechanical Turk (39.1% female, mean age of 37.2 years, ranging from 18 to 83 years; see Supplementary Table 1 for a full sample description). We collected data on 6 time-points in 1-week intervals between March 2nd and May 23rd and on August 27th and September 3rd, 2020. Data were always collected during U.S. working hours (beginning from about 8 a.m. to about 2 p.m. and lasting no later than 5 p.m.) in order to control for possible effects of weekday and time of day on the participant pool<sup>71</sup>. Participants provided informed consent prior to participating in the study. Checks of plausibility and completion time, as well as self-reports of data quality were used to screen out participants who did not pay sufficient attention. We excluded the data from a total of 167 participants due to missing data, missed attention checks, or text-field entries suggesting automated responses. This resulted in a total sample of  $N = 2296$  participants (or about 1362 participants in the withholding condition across timepoints, plus about 460 participants in each withdrawal condition at timepoints 3 and 4). The study adhered to the Declaration of Helsinki, relevant laws, and institutional guidelines, as certified by IRB of the University of Konstanz. The University of Konstanz's IRB approved the study. All participants provided informed consent.

## Design

Participants made a total of 19 hypothetical decisions to which of two patients a doctor should allocate the only available ventilator during a viral epidemic that had created a scarcity of medical resources. They could either decide to allocate the ventilator to one of the two patients or decide to use a randomization device (like a random lottery draw or fair coin flip). The two patients in each of the scenarios were described to differ in exactly one focal feature (implying no difference in survival probability). This allowed us to address the question how different patient features affect participants' allocation decisions, without introducing noise through interaction effects or potential diversions of richer representations. Based on our expectations of who will be favored due to their feature (e.g., based on higher contributions to society) we distinguish between favored and disfavored patients in the following. For a list of patient pairs see Fig. 1 (pairs are aligned so that the individual we expected to be saved with a higher probability is placed on the right). The instructions clearly stated that the patient who received the ventilator would be saved, whereas the other patient would die. In time-points 3 and 4 we additionally introduced a withdrawal condition and varied between-subjects whether the favored patient, the disfavored patient, or neither patient had been pre-allocated the last available ventilator. Thus, the decision became whether to retain the default allocation, withdraw and re-allocate the ventilator to the other patient, or opt for re-allocation by means of a randomization device. An example scenario can be found in the supplement (see Supplementary Fig. 1).

## Procedure

Participants first provided informed consent, received instructions that they will be asked to make decisions about which one of two comparable patients differing in one key regard should receive life-saving treatment. Patient pairs were presented in a randomized order. Next, participants answered questions about their own risk and their severity perceptions of the current COVID-19 crisis, their own health, altruism, time and risk preferences, as well as demographic questions, including their state and county, their political views and religiosity. To allow for the detection of in-group effects, we asked participants to provide information on characteristics equal or related to the characteristics of our hypothetical patients (e.g., political orientation and religious beliefs). Moreover, they provided information on their county so that we were able to assess the objective severity of the COVID-19 crisis in their area of residence. Participants always had the option to leave questions blank.

## Data analyses

Our two main outcomes were (a) the proportion of random allocations, with lower proportions indicating less impartiality patient feature, and (b) the proportion of allocations to the patient whom we expected to be favored among non-random allocations, with higher proportions indicating a bias towards the corresponding patient feature. Unless an analysis explicitly refers to a withholding condition, analyses are based on the withholding condition. None of the analyses were pre-registered. To compare allocation probabilities within patient features, we used binomial tests against 50% on either allocation alternative. In order to control for variation between participants and patient features, all other analyses used pairs of logistic mixed models (using `lme4`) for allocations to the patient who was expected to be favored (or keeping, see below) among non-random allocations (denoted by  $P(\text{favored}|\text{non-random})$  in the formulas below) and random allocations. For allocations to favored patients, we included the proportion of non-random allocations as an additional predictor. Across these models, we report Wald- $\chi^2$  tests (indicating the relevance of predictors) and calculated comparisons of interest from post hoc tests (using `emmeans`), correcting for multiple tests with a Sidak correction. All discrete variables were effect-coded and all continuous variables were z-standardized, with the exceptions of age, which was mean-centered. The first pair of models compared allocation patterns between patient features with fixed effects for patient features and time point of data collection, including a random effect for the participant, capturing unexplained variance between individuals (e.g., though differences in randomness aversion).

Model of bias (including the probability of random responses for participants and features):

$$P(\text{favored}|\text{non-random}) \sim b_{\text{feature}} + b_{\text{timepoint}} + p(\text{non-random}|\text{participant}) + p(\text{non-random}|\text{feature}) + \sigma_{\text{participant}}$$

Model of impartiality:

$$P(\text{random}) \sim b_{\text{feature}} + b_{\text{timepoint}} + \sigma_{\text{participant}}$$

Models describing our default conditions included random effects for both patient features and participants. The second pair of models compared the conditions of withholding and withdrawal for random allocations and the probability that the patient who already received care could keep the ventilator, including fixed effects for condition (withholding, withdrawing from favored, withdrawing from unfavored) and timepoint of data collection. Since situations of withholding lacked a patient receiving care, we used the probability that the expected to be favored patient received the ventilator as a comparison benchmark.

Model of keeping the ventilator:

$$P(\text{keep}) \sim b_{\text{default condition}} + b_{\text{timepoints}} + \sigma_{\text{participant}} + \sigma_{\text{feature}}$$

Model of random allocations in situations of withdrawal:

$$P(\text{random}) \sim b_{\text{default condition}} + b_{\text{timepoints}} + \sigma_{\text{participant}} + \sigma_{\text{feature}}$$

## Data availability

All data, code, and materials have been made publicly available at OSF and can be accessed at <https://osf.io/pn239/>.

Received: 14 September 2023; Accepted: 19 December 2023

Published online: 12 January 2024

## References

1. Emanuel, E. J. *et al.* Fair allocation of scarce medical resources in the time of Covid-19. *N. Engl. J. Med.* **382**, 2049–2055. <https://doi.org/10.1056/NEJMs2005114> (2020).
2. White, D. B. & Lo, B. A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA* **323**, 1773–1774. <https://doi.org/10.1001/jama.2020.5046> (2020).
3. Iserson, K. V. & Moskop, J. C. Triage in medicine, part I: Concept, history, and types. *Ann. Emerg. Med.* **49**, 275–281. <https://doi.org/10.1016/j.annemergmed.2006.05.019> (2007).
4. Bauman, C. W., McGraw, A. P., Bartels, D. M. & Warren, C. Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Soc. Pers. Psychol. Compass* **8**, 536–554. <https://doi.org/10.1111/spc3.12131> (2014).
5. Bostyn, D. H., Sevenhant, S. & Roets, A. Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychol. Sci.* **29**, 1084–1093. <https://doi.org/10.1177/0956797617752640> (2018).
6. Camerer, C. & Mobbs, D. Differences in behavior and brain activity during hypothetical and real choices. *Trends Cogn. Sci.* **21**, 46–56. <https://doi.org/10.1016/j.tics.2016.11.001> (2017).
7. FeldmanHall, O. *et al.* What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition* **123**, 434–441. <https://doi.org/10.1016/j.cognition.2012.02.001> (2012).
8. Foot, P. The problem of abortion and the doctrine of double effect. *Oxf. Rev.* **5**, 5–15 (1967).
9. Thomson, J. J. The trolley problem. *Yale Law J.* **94**, 1395–1415. <https://doi.org/10.2307/796133> (1985).
10. Persad, G., Wertheimer, A. & Emanuel, E. J. Principles for allocation of scarce medical interventions. *Lancet* **373**, 423–431. [https://doi.org/10.1016/S0140-6736\(09\)60137-9](https://doi.org/10.1016/S0140-6736(09)60137-9) (2009).
11. Dufner, A. Withdrawal of intensive care during times of severe scarcity: Triage during a pandemic only upon arrival or with the inclusion of patients who are already under treatment?. *Bioethics* **35**, 118–124. <https://doi.org/10.1111/bioe.12837> (2021).
12. Buckwalter, W. & Peterson, A. Public attitudes toward allocating scarce resources in the COVID-19 pandemic. *PLoS ONE* **15**, 1–20. <https://doi.org/10.1371/journal.pone.0240651> (2020).
13. Chan, L. *et al.* Which features of patients are morally relevant in ventilator triage? A survey of the UK public. *BMC Med. Ethics* **23**, 1–14. <https://doi.org/10.1186/s12910-022-00773-0> (2022).
14. Dowling, A., Lane, H. & Haines, T. Community preferences for the allocation of scarce healthcare resources during the COVID-19 pandemic: A review of the literature. *Public Health* **209**, 75–81. <https://doi.org/10.1016/j.puhe.2022.06.004> (2022).
15. Fallucchi, F., Faravelli, M. & Quercia, S. Fair allocation of scarce medical resources in the time of COVID-19: What do people think?. *J. Med. Ethics* **47**, 3–6. <https://doi.org/10.1136/medethics-2020-106524> (2020).
16. Jin, L., Huang, Y., Liang, Y. & Zhang, Q. Who gets the ventilator? Moral decision making regarding medical resource allocation in a pandemic. *J. Assoc. Consumer Res.* **6**, 159–167. <https://doi.org/10.1086/711734> (2021).
17. Lee, J.-S., Kim, S. & Do, Y. K. Public preferences for allocation principles for scarce medical resources in the COVID-19 pandemic in Korea: Comparisons with ethicists' recommendations. *J. Prev. Med. Public Health* **54**, 360–369. <https://doi.org/10.3961/jpmph.21.333> (2021).
18. Reeskens, T., Roosma, F. & Wanders, E. The perceived deservingness of COVID-19 healthcare in the Netherlands: A conjoint experiment on priority access to intensive care and vaccination. *BMC Public Health* **21**, 1–8. <https://doi.org/10.1186/s12889-021-10488-3> (2021).
19. Wilkinson, D., Zohny, H., Kappes, A., Sinnott-Armstrong, W. & Savulescu, J. Which factors should be included in triage? An online survey of the attitudes of the UK general public to pandemic triage dilemmas. *BMJ Open* <https://doi.org/10.1136/bmjopen-2020-045593> (2020).
20. Kahane, G. *et al.* Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychol. Rev.* **125**, 131–164. <https://doi.org/10.1037/rev0000093> (2018).
21. Fazio, R. H. & Olson, M. A. Implicit measures in social cognition research: Their meaning and use. *Annu. Rev. Psychol.* **54**, 297–327. <https://doi.org/10.1146/annurev.psych.54.101601.145225> (2003).
22. Gawronski, B., De Houwer, J. & Sherman, J. W. Twenty-five years of research using implicit measures. *Soc. Cogn.* **38**, s1–s25. <https://doi.org/10.1521/soco.2020.38.suppl.s1> (2020).
23. Cushman, F., Young, L. & Hauser, M. The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychol. Sci.* **17**, 1082–1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x> (2006).
24. Goodwin, G. P. & Landy, J. F. Valuing different human lives. *J. Exp. Psychol. Gen.* **143**, 778. <https://doi.org/10.1037/a0032796> (2014).
25. Waldmann, M. R., Nagel, J. & Wiegmann, A. Moral judgment. In *The Oxford Handbook of Thinking and Reasoning* 364–389 (Oxford University Press, 2012).
26. Dolan, P., Shaw, R., Tsuchiya, A. & Williams, A. QALY maximisation and people's preferences: A methodological review of the literature. *Health Econ.* **14**, 197–208. <https://doi.org/10.1002/hec.924> (2005).
27. Li, M., Vietri, J., Galvani, A. P. & Chapman, G. B. How do people value life?. *Psychol. Sci.* **21**, 163–167. <https://doi.org/10.1177/0956797609357707> (2010).
28. Werner, P. & Landau, R. Laypersons' priority-setting preferences for allocating a COVID-19 patient to a ventilator: Does a diagnosis of Alzheimer's disease matter?. *Clin. Interv. Aging* **15**, 2407–2414 (2020).
29. Ringel, M. M. & Ditto, P. H. The moralization of obesity. *Soc. Sci. Med.* **237**, 112399. <https://doi.org/10.1016/j.socscimed.2019.112399> (2019).
30. Heusler, A., Osiander, C. & Schmidtke, J. Essential for society but not equally deserving of preferential treatment? A discrete-choice experiment regarding COVID-19 healthcare. *Soc. Sci. Med.* **311**, 115304. <https://doi.org/10.1016/j.socscimed.2022.115304> (2022).
31. van Oorschot, W. Who should get what, and why? On deservingness criteria and the conditionality of solidarity among the public. *Policy Polit.* **28**, 33–48. <https://doi.org/10.1332/0305573002500811> (2000).
32. Sprengholz, P., Korn, L., Felgendreiff, L., Eitze, S. & Betsch, C. A lay perspective on prioritization for intensive care in pandemic times: Vaccination status matters. *Clin. Ethics* <https://doi.org/10.1177/1477509221094474> (2022).
33. Krupp, D. B., DeBruine, L. M. & Barclay, P. A cue of kinship promotes cooperation for the public good. *Evol. Hum. Behav.* **29**, 49–55. <https://doi.org/10.1016/j.evolhumbehav.2007.08.002> (2008).
34. Everett, J. A. C., Faber, N. S. & Crockett, M. Preferences and beliefs in ingroup favoritism. *Front. Behav. Neurosci.* **9**, 15. <https://doi.org/10.3389/fnbeh.2015.00015> (2015).
35. Lane, T. Discrimination in the laboratory: A meta-analysis of economics experiments. *Eur. Econ. Rev.* **90**, 375–402. <https://doi.org/10.1016/j.euroecorev.2015.11.011> (2016).
36. Tajfel, H. Experiments in intergroup discrimination. *Sci. Am.* **223**, 96–103 (1970).

37. Tajfel, H., Billig, M. G., Bundy, R. P. & Flament, C. Social categorization and intergroup behaviour. *Eur. J. Soc. Psychol.* **1**, 149–178. <https://doi.org/10.1002/ejsp.2420010202> (1971).
38. Yamagishi, T. & Kiyonari, T. The group as the container of generalized reciprocity. *Soc. Psychol. Q.* **63**, 116–132. <https://doi.org/10.2307/2695887> (2000).
39. Bleske-Rechek, A., Nelson, L. A., Baker, J. P., Remiker, M. W. & Brandt, S. J. Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a romantic partner. *J. Soc. Evol. Cult. Psychol.* **4**, 115. <https://doi.org/10.1037/h0099295> (2010).
40. Christensen, J. F. & Gomila, A. Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neurosci. Biobehav. Rev.* **36**, 1249–1264. <https://doi.org/10.1016/j.neubiorev.2012.02.008> (2012).
41. Cikara, M., Farnsworth, R. A., Harris, L. T. & Fiske, S. T. On the wrong side of the trolley track: Neural correlates of relative social valuation. *Soc. Cogn. Affect. Neurosci.* **5**, 404–413. <https://doi.org/10.1093/scan/nsq011> (2010).
42. Petrinovich, L., O'Neill, P. & Jorgensen, M. An empirical study of moral intuitions: Toward an evolutionary ethics. *J. Pers. Soc. Psychol.* **64**, 467–478. <https://doi.org/10.1037/0022-3514.64.3.467> (1993).
43. Rai, T. S. & Fiske, A. P. Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychol. Rev.* **118**, 57–75. <https://doi.org/10.1037/a0021867> (2011).
44. Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D. & Ditto, P. H. The motivated use of moral principles. *Judgm. Decis. Mak.* **4**, 479–491. <https://doi.org/10.1017/S1930297500004022> (2009).
45. Awad, E. *et al.* The moral machine experiment. *Nature* **563**, 59–64. <https://doi.org/10.1038/s41586-018-0637-6> (2018).
46. Lenton, A. P., Blair, I. V. & Hastie, R. The influence of social categories and patient responsibility on health care allocation decisions: Bias or fairness?. *Basic Appl. Soc. Psychol.* **28**, 27–36. [https://doi.org/10.1207/s15324834basps2801\\_3](https://doi.org/10.1207/s15324834basps2801_3) (2006).
47. Helbling, M., Maxwell, R., Munzert, S. & Traunmüller, R. The importance of citizenship for deserving COVID-19 treatment. *Humanit. Soc. Sci. Commun.* **9**, 1–8. <https://doi.org/10.1057/s41599-022-01311-4> (2022).
48. Larsen, M. H. & Schaeffer, M. Healthcare chauvinism during the COVID-19 pandemic. *J. Ethn. Migr. Stud.* **47**, 1455–1473. <https://doi.org/10.1080/1369183X.2020.1860742> (2021).
49. Stoetzer, L. F. *et al.* Affective partisan polarization and moral dilemmas during the COVID-19 pandemic. *Polit. Sci. Res. Methods* **11**, 429–436. <https://doi.org/10.1017/psrm.2022.13> (2022).
50. Greenwald, A. G. & Banaji, M. R. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol. Rev.* **102**, 4–27. <https://doi.org/10.1037/0033-295X.102.1.4> (1995).
51. Jost, J. T. *et al.* The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Res. Organ. Behav.* **29**, 39–69. <https://doi.org/10.1016/j.riob.2009.10.001> (2009).
52. Nosek, B. A., Banaji, M. R. & Greenwald, A. G. Harvesting intergroup attitudes and stereotypes from a demonstration website. *Group Dyn. Theory Res. Pract.* **6**, 101–115. <https://doi.org/10.1037/1089-2699.6.1.101> (2002).
53. Stanley, D. A., Sokol-Hessner, P., Banaji, M. R. & Phelps, E. A. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proc. Natl. Acad. Sci.* **108**, 7710–7715. <https://doi.org/10.1073/pnas.1014345108> (2011).
54. De Dreu, C. K. W., Greer, L. L., Kleef, G. A. V., Shalvi, S. & Handgraaf, M. J. J. Oxytocin promotes human ethnocentrism. *Proc. Natl. Acad. Sci.* **108**, 1262–1266. <https://doi.org/10.1073/pnas.1015316108> (2011).
55. Pearson, A. R., Dovidio, J. F. & Gaertner, S. L. The nature of contemporary prejudice: Insights from aversive racism. *Soc. Pers. Psychol. Compass* **3**, 314–338. <https://doi.org/10.1111/j.1751-9004.2009.00183.x> (2009).
56. Chen, Y. & Li, S. X. Group identity and social preferences. *Am. Econ. Rev.* **99**, 431–457. <https://doi.org/10.1257/aer.99.1.431> (2009).
57. Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R. & Phelps, E. A. The price of racial bias: Intergroup negotiations in the ultimatum game. *Psychol. Sci.* **24**, 2498–2504. <https://doi.org/10.1177/0956797613496435> (2013).
58. Banaji, M. R., Bhaskar, R. & Brownstein, M. When bias is implicit, how might we think about repairing harm?. *Curr. Opin. Psychol.* **6**, 183–188. <https://doi.org/10.1016/j.copsyc.2015.08.017> (2015).
59. Dwenger, N., Kübler, D. & Weizsäcker, G. *Flipping a Coin: Theory and Evidence*. Working Paper No. 2012–14, (Max Planck Institute for Tax Law and Public Finance, 2012).
60. Dana, J., Weber, R. A. & Kuang, J. X. Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Econ. Theory* **33**, 67–80. <https://doi.org/10.1007/s00199-006-0153-z> (2007).
61. Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H. & Strongman, J. A. Moral hypocrisy: Appearing moral to oneself without being so. *J. Pers. Soc. Psychol.* **77**, 525–537. <https://doi.org/10.1037/0022-3514.77.3.525> (1999).
62. Diekmann, K. A., Samuels, S. M., Ross, L. & Bazerman, M. H. Self-interest and fairness in problems of resource allocation: Allocators versus recipients. *J. Pers. Soc. Psychol.* **72**, 1061–1074. <https://doi.org/10.1037/0022-3514.72.5.1061> (1997).
63. Baron, J. & Ritov, I. Reference points and omission bias. *Organ. Behav. Hum. Decis. Process.* **59**, 475–498. <https://doi.org/10.1006/obhd.1994.1070> (1994).
64. Samuelson, W. & Zeckhauser, R. Status quo bias in decision making. *J. Risk Uncertain.* **1**, 7–59. <https://doi.org/10.1007/BF00055564> (1988).
65. Norman, R. *et al.* Public preferences for allocating ventilators in an intensive care unit: A discrete choice experiment. *Patient-Patient-Centered Outcomes Res.* **14**, 319–330. <https://doi.org/10.1007/s40271-021-00498-z> (2021).
66. Street, A. E., Street, D. J. & Flynn, G. M. Who gets the last bed? a discrete-choice experiment examining general population preferences for intensive care bed prioritization in a pandemic. *Med. Decis. Mak.* **41**, 408–418. <https://doi.org/10.1177/0272989X21996615> (2021).
67. Keren, G. & Teigen, K. H. Decisions by coin toss: Inappropriate but fair. *Judgm. Decis. Mak.* **5**, 83–101. <https://doi.org/10.1017/S1930297500000942> (2010).
68. Mummolo, J. & Peterson, E. Demand effects in survey experiments: An empirical assessment. *Am. Polit. Sci. Rev.* **113**, 517–529. <https://doi.org/10.1017/S0003055418000837> (2019).
69. Conway, P., Goldstein-Greenwood, J., Polacek, D. & Greene, J. D. Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition* **179**, 241–265. <https://doi.org/10.1016/j.cognition.2018.04.018> (2018).
70. Everett, J. A. C. & Kahane, G. Switching tracks? Towards a multidimensional model of utilitarian psychology. *Trends Cogn. Sci.* **24**, 124–134. <https://doi.org/10.1016/j.tics.2019.11.012> (2020).
71. Arechar, A. A., Kraft-Todd, G. T. & Rand, D. G. Turking overtime: How participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. *J. Econ. Sci. Assoc.* **3**, 1–11. <https://doi.org/10.1007/s40881-017-0035-0> (2017).

# Acknowledgements

This research was funded by the University of Konstanz. N.G., W.G., H.N. and H.G. received additional funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2117–422037984. W.G., H.G., and N.G. received additional funding from DFG Grant Number 441541975.

### Author contributions

N.G., H.N. and W.G. contributed to the conceptual development of the work and designed the experiments; N.G. programmed the experiments, collected the data, and analyzed the data with input from H.N., W.G., and H.G.; N.G. produced the first draft, all authors provided critical revisions, and approved of the final version.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-50385-w>.

**Correspondence** and requests for materials should be addressed to N.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024