

When waste pays:

Inefficient (but seemingly fair) resource allocators receive social and economic rewards

Bradley R. DeWees
Harvard University and United States Air Force

Charles A. Dorison
Harvard University

Zoe Rahwan
Max Planck Institute for Human Development

Christopher Robichaud
Harvard University

Jennifer S. Lerner
Harvard University

Abstract

The tension between equality and efficiency presents a cardinal trade-off in scarce resource allocation decisions. Three pre-registered experiments (N=1,095) and an independent replication (N=300) drew on the revised value pluralism model (Tetlock, Peterson, & Lerner, 1996) to predict how decision makers resolve such trade-offs. Studies 1-2 found that social observation by non-stakeholders increased allocators' preferences for equal (yet inefficient) allocations – a finding that held even with real money and even when the efficient allocation made one party better off and no one worse off. Study 3 found that allocators who made inefficient choices received positive evaluations and monetary rewards from observers. Analyses also examined the boundary conditions for such benefits and the underlying mechanisms. Taken together, the results elucidate causal mechanisms for a fact of political life: Decision makers who make equal rather than efficient allocations can, by virtue of doing so, receive greater financial and social rewards from observers.

Keywords: decision making, efficiency, equality, person perception, open science

WHEN WASTE PAYS

“Liberte, egalité, fraternite!” So chanted French revolutionaries in the late 1700s. “Equal rights!” contemporary crowds chant. “Freedom for all!” they cheer. Praise for values is often absolute. Rare is the occasion when someone acknowledges that values, no matter how cherished in isolation, frequently come into tension with one another. A crowd would not proclaim: “We support equality unless it comes at the cost of efficiency.”¹

Indeed, sharp resistance might arise if someone were asked to quantify exactly how much equality they were willing to sacrifice in order to increase efficiency (Tetlock et al., 2000). The resistance might be especially strong if the trade-offs were made under the watchful eyes of observers (Lerner & Tetlock, 1999). And yet, such trade-off reasoning is essential to rational decision making (Becker, 1981) because values do inevitably clash in real and important ways. A hospital can, for example, allocate kidneys either to patients with the highest chance of survival (maximizing efficiency) or to all patients equally, regardless of chances of survival (maximizing equality). The present paper examines such value conflicts, elucidating how and when the presence of an evaluative audience may alter trade-off reasoning. In particular, we examine decision processes when *maximizing equality conflicts with maximizing efficiency* (cf. Choshen-Hillel, Shaw, & Caruso, 2015).

The Value Pluralism Model

Drawing on decades of research on human values (e.g., Inglehart & Baker, 2000; Schwartz, 1992; Rokeach, 1973), the value pluralism model (VPM; Tetlock, 1986) provides a theoretical starting point for predicting how individuals manage value trade-offs. Contrary to the view that different values and principles are reducible to one supervalue or single moral principle, e.g., Kant’s Categorical Imperative (Kant, 1797/1948), a core premise of the VPM is

¹ Efficiency in this sense means not wasting resources (see Gordon-Hecker, Choshen-Hillel, Shalvi, & Bereby-Meyer, 2017).

WHEN WASTE PAYS

that individuals possess an implicit internal hierarchy of competing core values. Moreover, the ordinal ranking of such values guides decisions requiring trade-offs among values (Suedfeld, Bluck, Loewen, & Elkins, 1994; Tetlock, 1986; Tetlock, Michelitti, & Hanum, 1984).

An important revision of the original model, the revised value pluralism model (RVPM: Tetlock, Peterson, & Lerner, 1996), explicitly sought to tighten the model's specificity and broaden its applicability by incorporating two key postulates. One postulate addressed the ways in which *the social content of colliding values* would alter trade-off reasoning. It predicted, for example, that decision makers would find it so painful to make trade-offs between a sacred value and a secular value (e.g., money versus freedom – a taboo trade-off; Tetlock et al., 2000), they might denigrate anyone who suggests they do so. Considerable evidence to date supports this prediction as well as the postulate more generally (e.g., Tetlock et al., 2000; Lichtenstein, Gregory, & Irwin, 2007; Fiske & Tetlock, 1997).

The RVPM's second postulate, examining the *social context of colliding values*, has, by contrast, yet to be rigorously tested. It holds that (a) when making trade-offs of real consequence, individuals are implicitly or explicitly accountable to important constituent audiences and (b) the presence of such audiences will systematically shape how decision makers trade off competing values (Tetlock, Peterson, & Lerner, 1996). The present studies aim to test this *social context* postulate, examining whether observation by an audience will systematically alter how decision makers resolve equality-efficiency trade-offs.

To do so, we draw on Adelberg and Batson's (1978) clever resource-allocation paradigm. In their studies, hypothetical financial-aid agents were either unaccountable or accountable for their resource allocations to key stakeholders. When hypothetical financial resources were inadequate to cover the educational needs of all potential recipients (each needed a certain

WHEN WASTE PAYS

amount to enroll for the semester; less than that amount was stipulated to be of no use), results revealed that *only* unaccountable agents matched awards to needs effectively (Adelberg & Batson, 1978), specifically by giving enough to some applicants (efficient strategy) rather than a little to all applicants (equality strategy). Thus, accountability to stakeholders caused decision makers to be inefficient but “fair.”² Although mechanisms were not assessed, it may be that pressure from stakeholders led accountable allocators to believe that treating all potential recipients equally (an apparent fairness heuristic) was the best strategy for appearing trustworthy – a trait that observers want key decision makers to possess (e.g., Rousseau, Sitkin, Burt, & Camerer, 1998; Fiske & Dupree, 2014).

Integrating predictions from the RVPM with Adelberg and Batson’s findings (1978), Study 1 sought to examine whether being evaluated, even by someone without a stake in the outcome – i.e., someone interested only in whether one is a trustworthy person generally, would be sufficient to trigger inefficient allocations. We term this the *wasteful-when-watched hypothesis*. Study 2 tested the same hypothesis but with a different allocation domain, a real evaluative audience, and real financial stakes. Study 3 turned from assessing the behavior of allocators to assessing the behavior of observers, examining the hypothesis that allocators’ inefficient behavior may be inefficient in economic terms yet simultaneously socially rewarding – the *waste-pays hypothesis*. Analyses also examined the boundary conditions for such benefits and underlying mechanisms. In sum, the studies aimed to investigate whether (a) the mere presence of an evaluative audience would shift allocators toward equality and away from efficiency and (b) observers would reward allocators who exhibit such shifts. This pattern, if upheld, would hold implications not only for understanding the psychology of value tradeoffs but

² In a related and important line of work, though one that did not directly examine the role of an accounting audience, Chosen-Hillel et al. (2015) found that social concerns increase equality-favoring choices.

WHEN WASTE PAYS

also for the design of any allocation systems that seek to make maximal utility of scarce resources.

Study 1: Scarce Resource Allocation with Hypothetical Organs

Method

Study 1 examined the *wasteful-when-watched hypothesis* – i.e., that the mere prospect of being evaluated as a trustworthy decision maker would lead participants to make more equal, yet inefficient, allocations of scarce resources. It also examined whether relatively heightened thoughts about equality, as well as diminished thoughts about efficiency, would mediate such an effect.

Open science. In keeping with best practices for fully reproducible science (Simmons, Nelson, & Simonsohn, 2011), for this and all subsequent studies we pre-registered all hypotheses and methodological decisions (e.g., determining sample size). We also made available all preregistrations, manipulations, measures, data, code, and materials available via an Open Science Framework (OSF) repository [here](#).

Participants. Based on simulations from pilot data, we determined that a sample size of 425 participants would yield adequate power to test our primary hypothesis. We increased the estimate by 25 participants per cell to account for preregistered exclusions (of participants who failed to follow writing instructions). This yielded a target sample size of 475. After advertising an “Ethical Decision Making Survey” via Amazon mTurk, we collected 480 complete survey responses. Thirty participants did not follow instructions and were removed. This resulted in a final sample of 450 ($M_{age} = 37.3$, 44.4% female). Importantly, robustness checks on our data reveal that the overall results remain consistent with the patterns reported below regardless of whether we exclude those who failed to follow instructions (see OSF page for data).

WHEN WASTE PAYS

Design. All participants, hereafter called “allocators,” made hypothetical allocations of life-saving organs. Before they did so, we randomly assigned allocators either to consider how the general public would evaluate them as a trustworthy decision maker (Evaluation condition) or to privately consider what they themselves deemed to be an important decision-making factor (Control condition). We counter-balanced the order in which the equal and efficient allocation options appeared to decision makers. The two factors resulted in a 2 (Focus: Evaluation, Control) X 2 (Order: Equal-first, Efficient-first) between-subjects design.

Procedure. Adapting a scenario from Ubel and Loewenstein (1996), we asked all allocators to make an organ allocation decision as the director of a regional allocation board in the United States. Introducing a situation of scarcity, we instructed allocators that 200 children were on a waiting list to receive a liver transplant but that only 100 livers were available for distribution. The children were organized into two groups of 100 according to their likelihood of survival after a liver transplant. In one group of 100 children, each child had an 80% chance of survival (high expected-value group); in the other group of 100 children, each child had a 50% chance of survival (low expected-value group). The participant’s task was to “decide what percentage of the 100 livers should go to each group.” A concern for efficiency in this scenario would lead allocators to distribute the livers to the group with the higher likelihood of survival, leading to a higher expected number of survivors for a given allocation of resources. A concern for equality would lead allocators to distribute the livers equally to each group, giving all children an equal chance at receiving a liver but leading to a lower number of expected survivors. Our primary dependent variable was the proportion of livers allocated to the group with the higher likelihood of survival.

WHEN WASTE PAYS

Before making their choices, all participants read the following: “In a difficult decision like this, it is possible to consider many different factors, all of which are important.” Then they received one of two instructions according to condition. Those in the *Evaluation condition* read: “We would urge you to consider the factor of trustworthiness in the eyes of the general public.” Those in the *Control condition* read: “We would urge you to consider the factor that you believe is most important.”

Instructions cued participants to write a minimum of 150 characters about their choice. This writing task allowed us to test the extent to which relative attentional focus from efficiency to equality would underpin the potential effect on inefficient allocations of the scarce resources.

Dependent variable. After writing, allocators made their allocation decision by indicating the percentage of livers they would like to send to each group using a sliding bar (the survey software forced their allocations across the two groups to sum to 100). Again, our dependent variable was the percentage of livers allocated to the group with the higher likelihood of survival.

Exploratory variables. Before concluding, allocators had the opportunity to provide any other rationale they wished to express for their choice. Finally, the study concluded with a series of exploratory items concerning demographics, political ideology, education, whether religion was important in their life, gender, and age. See supplementary analyses on OSF for these results.

Results

Manipulation checks. Two coders who were blind to condition and allocation choice used a binary measure to indicate whether a participant expressed a concern in their written responses for gaining the public’s trust (1 if yes, 0 if no). The two raters agreed 97% of the time

WHEN WASTE PAYS

and showed a Cohen's kappa of .93. In the cases of disagreement, we averaged the two raters' scores; possible variable levels were thus 0, 0.5, and 1. Regressing this variable on Focus condition (Evaluation vs. Control) showed an effective manipulation (95% CI [0.45, 0.58], $t = 15.98$, $p < .001$).

As expected, we observed neither a main effect of Order (95% CI [-5.02, 3.27], $t = -.416$, $p = .67$) nor an interaction between Order and Condition (95% CI [-6.57, 9.96], $t = .403$, $p = .69$). Thus, we collapsed across the Order variable in all ensuing analyses.

Main effect. To determine whether the Evaluation manipulation would trigger inefficient allocations (i.e., support for the *wasteful-when-watched hypothesis*), we compared the proportion of allocations to the efficient group across conditions. Consistent with the hypothesis, decision makers in the Control condition allocated 78.5% of their organs to the group with the highest chance of survival, while decision makers in the Evaluation condition allocated just 73.5% to the high-survival group (95% CI [-9.11, -0.87]; *cohen's d* = 0.23; $t = -2.38$; $p = .018$).

Replication of Study 1. We ran an initial, lean version of Study 1 without pre-registration, which allowed for a direct replication of the results reported above. Using a slightly smaller sample ($N = 300$ based on simulated power analysis), we found the same pattern of results: Those in the Evaluation condition again made less efficient allocations than those in the Control condition (95% CI [-10.32, -0.99]; *cohen's d* = 0.28; $t = -2.38$; $p = .018$).

Mediation analysis. Although the main result is consistent with the RVPM prediction (that the nature of the social context would influence the relative attentional priority given to one value over another), this choice outcome in and of itself cannot illuminate the underlying mechanisms. To gain insight into the mechanism, we conducted a mediation analysis using the written responses of allocators. We predicted that evaluation would shift attentional focus from

WHEN WASTE PAYS

equality to efficiency and that this shift in attention focus would be associated with resource allocation decisions.

Two research assistants, blind to hypotheses and treatment conditions, coded participant writings to determine the extent to which they focused on 1) efficiency and 2) equality. We used the same procedure for resolving any coder disagreements as we used when resolving disagreements on the manipulation check data. Having observed acceptable Cohen's kappas of .78 (91% agreement) and .86 (95% agreement) for efficiency and equality, respectively, we fit a structural equation model using the Lavaan package in R (Rosseel, 2011); the results appear in Figure 1.

As compared to allocators in the Control condition, allocators in the Evaluation condition were *more likely* to focus on equality (95% CI [0.05, 0.20]; $t = 3.31$; $p = .001$) and *less likely* to focus on efficiency (95% CI [-0.33, -0.18]; $t = -6.68$; $p < .001$). While a focus on equality was associated with *less* efficient allocations (95% CI [-18.45, -9.83]; $t = -6.43$; $p < .001$), a focus on efficiency was associated with *more* efficient allocations (95% CI [17.05, 25.57]; $t = 9.81$; $p < .001$). Taken together, we observed a significant indirect effect of condition on inefficient resource allocations through both an increased focus on equality-related concerns (95% CI [-2.95, -0.59]; $t = -2.95$; $p = .003$) and a decreased focus on efficiency-related concerns (95% CI [-7.39, -3.51]; $t = -5.52$; $p < .001$).

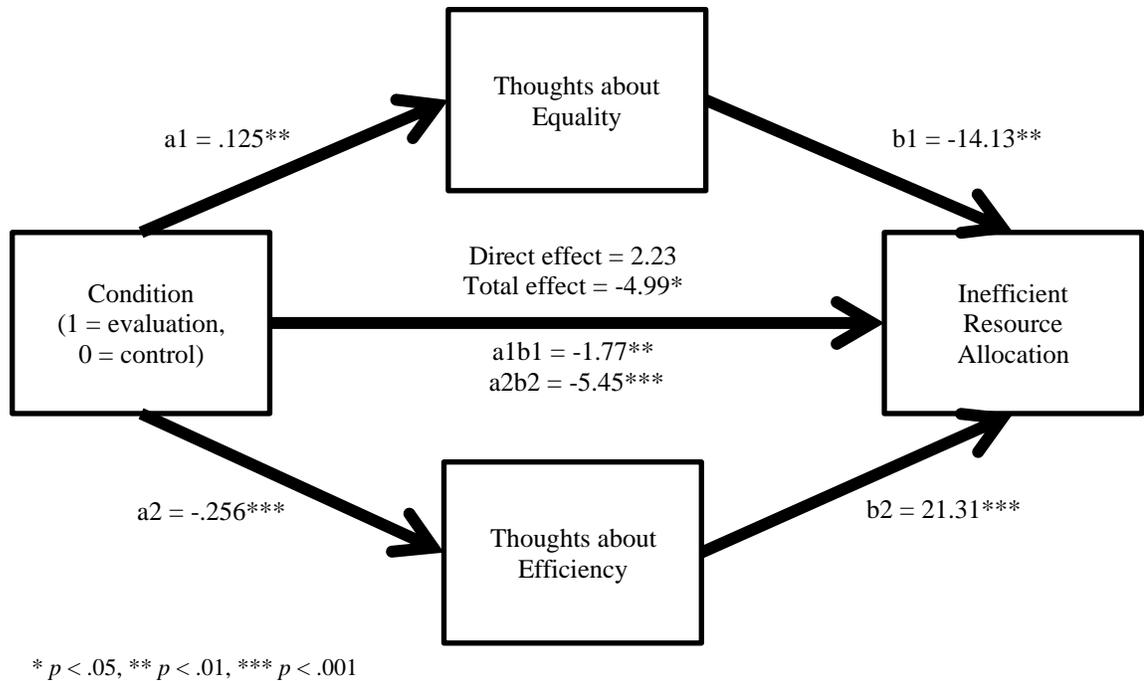


Fig. 1. Individuals in the evaluation condition made more inefficient resource allocations than individuals in the control condition. As judged by their written statements, they were more likely to focus on equality and less likely to focus on efficiency. Increased focused on equality and decreased focus on efficiency were in turn associated with inefficient resource allocations.

Discussion

Consistent with predictions from the RVPM, Study 1 found support for the *wasteful-when-watched hypothesis*: the prospect of public evaluation increased inefficient (albeit equal) allocations. Study 1 also provided evidence consistent with the hypothesis that heightened attentional focus on equality-related concerns, as well as diminished attentional focus on efficiency-related concerns, partially mediated this effect. Thus, the presence of an evaluative audience not only changed what decision makers ultimately chose (in this case, inefficiency) but also how decision makers thought about the choice.

Study 2: Scarce Resource Allocation with Real Financial Stakes

Method

WHEN WASTE PAYS

In order to test for the importance and generality of Study 1's results, Study 2 sought to test the *wasteful-when-watched hypothesis* with a different allocation domain, a real evaluative audience, and real financial stakes. In this case, participants chose between (a) an economically inefficient but equal allocation of real money or (b) an economically efficient but unequal allocation of real money. Participants (hereafter "allocators") were randomly assigned to make their allocation decision either while being observed by a future partner in an economic game (i.e., the Evaluation condition) or anonymously (i.e., the Control condition). We predicted that allocators in the Evaluation condition would be more likely to make inefficient choices than would participants in the Control condition.

Participants. Simulations from pilot data suggested that a sample size of 450 would adequately power tests of the hypothesis. We recruited 450 participants and (by chance factors in the online sign-up procedure) ended up with 457 participants via mTurk for a "Survey on Judgment and Decision-Making." To have real observers and to avoid deception, we recruited a similar number of participants to play in the opposite side of the economic game (explained in more detail below). Because these participants were not pertinent to our hypotheses, we describe that sample and their choices in supplementary material on OSF.

In line with the pre-registered analysis plan, and following standard practice for online samples, the final sample included only participants who paid attention to the instructions (indicated by correctly answering at least the pre-set threshold of five out of seven basic comprehension questions). In line with commonly observed rates, approximately 11% ($n = 53$) of participants failed to meet this criterion, resulting in a final sample of 404 ($M_{age} = 37.25$, 52.5% female). Importantly, robustness checks on our data reveal that the overall results remain consistent with the patterns reported below regardless of how we adjust this exclusion criterion.

WHEN WASTE PAYS

Design. Participants were randomly assigned to make their allocation decision while being observed by a future trust game partner or anonymously. As in Study 1, to avoid introducing confounds, we also randomly assigned the order in which the allocation options appeared. This yielded a 2 (Focus: Evaluation, Control) X 2 (Order of Allocation Options: Equal-first, Efficient-first) between-subjects design. The dependent variable was a binary choice between an economically inefficient (yet equal) allocation or an efficient (yet unequal) allocation. This choice had real financial stakes, as detailed below.

Procedure. Allocators read that they would learn about each of two tasks before engaging in the first task. In the first task, they would allocate bonus payments efficiently or equally to other mTurk workers (hereafter “recipients”). We generically described this efficient or equal allocation as a “dilemma.” They also read that, in the second task, they would play an economic game with another mTurk worker (hereafter a “partner”). We generically described this as a “game.” We clarified that the “recipients” and the “partner” were not the same for this first task.

Following standard procedures (Berg, Dickhaut, & McCabe, 1995; Jordan, Hoffman, Nowak, & Rand, 2016), the game instructions (for the second task) stated that another mTurker—the partner—would begin with \$.30 and would have the opportunity to transfer any amount, including none at all, to the allocator. We (the experimenters) would triple any amount partners transferred. Allocators would be able to decide how much of the tripled amount, if any, they wanted to transfer *back* to their partner.

Allocators in the *Evaluation condition* read that their partner would “decide whether to play the game with you based on your decision in the dilemma.” Allocators in the *Control condition* read that their partner would “have no knowledge of your decision in the dilemma.

WHEN WASTE PAYS

Your choice in the dilemma will NOT influence the outcome of the game.” The opportunity to play the game was an opportunity to be given more money; the existence of the game therefore created an incentive for allocators in the evaluation condition to make an allocation in the first task that would be likely to persuade partners to let them play the game.

Dependent variable. Allocators in both experimental conditions made one allocation decision, which served as the primary dependent variable. Allocators read that three other mTurkers—“recipients,” none of whom was the partner who would play in the trust game—had recently completed a task. Allocators could award these recipients up to four bonus tokens worth \$.20 each. They could divide the tokens in one of two ways. Allocators could award: (a) three of the four available tokens, giving one token to each recipient (equal allocation) or (b) all four tokens, giving one token to two recipients and two tokens to a third (unequal but efficient allocation). Thus, the efficient allocation was Pareto-optimal – i.e., it made one of the recipients better off while making none of them worse off (Hochman & Rodgers, 1969). The order in which these two allocation options appeared varied randomly according to the *Order condition*.

After this choice, and to avoid using deception, allocators then actually played the trust game by indicating what percentage, if any, they wanted to return to their partner. These data, while not of inferential interest, are available on OSF.

Exploratory variables. As in Study 1, they also completed a set of exploratory items. We sought to assess the extent to which the impact of the evaluation condition involved (a) purely strategic behavior driven by a desire to maximize the reward from the game by anticipating the views of the partner, or (b) a change in the participant’s own personal valuation of equality. We used a combination of approaches to explore these mechanisms, including

WHEN WASTE PAYS

measuring individual differences in strategic reasoning, probing allocator motives, and comparing allocator reciprocity across experimental conditions.

Manipulation check. Participants completed a manipulation check, which asked them to rate the extent to which they agreed with the following statement, if at all: “When deciding between the equal and efficient allocation, I considered how the Sender [the Partner] would view my choice.”

Results

Manipulation checks. Data indicated that the central manipulation succeeded: Evaluated allocators were more likely to agree with the statement that they considered how their trust game partner would view their allocation choice (95% CI [1.62, 2.37], $t = 10.46$, $p < .001$). We did not observe a main effect for Order on allocation decision (choice between equal or efficient allocation; *log-odds* 95% CI [.78, 1.70]; $t = -0.71$; $p = .479$). Contrary to predictions, however, and to the pattern in Study 1, this time we observed an interaction between the Evaluation condition and the Order in which we presented the equal and efficient allocations (*log-odds* 95% CI [1.17, 5.70]; $t = 2.34$; $p = .019$). Results across the three studies (Study 1 reported above, Study 1 replication available online, and Study 2) imply that the interaction with order is unreliable. For the sake of simplicity, we therefore report the results below collapsed across order.

Main effect. Our main hypothesis was that, even with real money on the line, allocators in the Evaluation condition would be more likely to make equal (yet inefficient) allocations than those in the Control condition. This was, in fact, the case. Supporting the *wasteful-when-watched hypothesis*, allocators in the Evaluation condition chose the efficient allocation 47% of the time, while allocators in the Control condition chose the efficient allocation 59% of the time (*log-odds*

WHEN WASTE PAYS

95% *CI* [.42, .92]; *cohen's d* = .24, *t* = -2.38; *p* = .017). Indeed, allocators in the Evaluation condition were approximately 20% ((59-47)/59) less likely to choose the efficient allocation than were allocators in the Control condition.

Exploratory analyses. Lending further support to the idea that presence of an evaluative audience truly increased preferences for equality, we conducted three exploratory sets of analyses. While full results are available on our OSF repository, we briefly summarize them here. First, we find evidence that observed allocators were marginally more likely to return more money in the trust game than their unobserved counterparts ($M_{Observed} = 36.87$, $M_{Anonymous} = 32.96$, 95% *CI* [-0.51, 8.33], *cohen's d* = .17, *t* = 1.74, *p* = .083), suggesting that observation did not simply increase strategizing in order to maximize selfish gains. Second, when asked which choice would convince the partner to send as much money as possible, most participants (51%) did not express a preference one way or the other (they believed that the two allocations would have similar effects). Those who did were nearly evenly divided over which allocation decision would be most effective (22% favored efficient, 27% favored equality; this latter division was indistinguishable from chance, $\chi^2 = .818$, *p* = .366). Finally, we did not observe an interaction between our measure of strategic reasoning and our Observed condition (*log-odds* 95% *CI* [0.34, 1.96]; *t* = -0.46; *p* = .646). High strategic reasoners—individuals most likely to anticipate the behavior or judgment of others and adapt their own behavior accordingly (LeVeck et. al., 2014)—were no more likely to change their allocation in response to observation. In short, we find no evidence that our effect is due purely to strategizing.

Discussion

Study 2 tested the *wasteful-when-watched hypothesis* in the presence of real financial stakes, a real evaluative audience, and a new domain of scarce resource. To create an especially

WHEN WASTE PAYS

conservative test, the economically efficient choice option introduced in Study 2 was also Pareto-optimal – it expanded the resource pie, making one party better off and no one worse off. Despite the Pareto-optimal alternative, the inefficient choice option became the preferred option when decision makers were observed by someone with whom they would later interact in an economic game.

Study 3: Social Perceptions of Allocators

The first two studies examined whether the presence of an evaluative audience would alter allocators' choices in an equality-efficiency trade-off. Our third study builds on the first two by asking a related question: Do resource allocators make a socially savvy choice when they uphold equality by sacrificing efficiency?

We based our hypotheses for this study on research examining how observers view decision makers who rely on deontological moral reasoning (i.e., reasoning based on the inherent morality of an action, regardless of outcomes), rather than consequentialist moral reasoning (i.e., reasoning aimed at maximizing the greatest good for the greatest number of people, regardless of means). Specifically, research finds that those who rely on deontological reasoning in sacrificial dilemmas are viewed as more morally trustworthy than those who rely on consequentialist reasoning (Everett, Pizarro, & Crockett, 2016; Everett, Faber, Savulescu, & Crockett, 2018; Sacco, Brown, Lustgraaf, & Hugenberg, 2017; Uhlmann, Zhu, & Tannenbaum, 2013).

Building off this work, we pre-registered three hypotheses. First, we predicted that, on average, observers would trust allocators who make equal allocations more than allocators who make efficient allocations, which we call the *waste-pays hypothesis*. Second, we hypothesized an interaction such that the effect of allocator choice on observer trust would depend on an observer's own equal-efficient preference. Specifically, we hypothesized that, while observers

WHEN WASTE PAYS

who prefer equality would reward allocators who choose equality (i.e., the *waste-pays hypothesis*), observers who prefer efficiency would show an attenuation of this effect. Drawing on the RVPM (Tetlock, Peterson, & Lerner, 1996), the rationale is that observers who value equality over efficiency should be more likely to treat that value as sacred – i.e., not a value one can trade off (Tetlock, Peterson, & Lerner, 1996).

Finally, we hypothesized that this asymmetric pattern of equality-preferring observers mistrusting those who disagree would be underpinned by an asymmetric pattern of warmth perceptions. Specifically, equality-preferring observers should perceive allocators who also prefer equality as warmer than allocators who prefer efficiency. By contrast, efficiency-preferring observers (pragmatic Utilitarians) should be less likely to make divergent attributions of warmth based on allocation behavior. We predicted that this asymmetry in perceptions of warmth would mediate the asymmetry in actual trusting behavior.

Method

Participants. Simulations with pilot data showed that samples of 200 or more would yield at least 80% power for all hypotheses related to observer behavior. To account for possible exclusions from non-compliance, we added 50 to this number for a sample goal of 250 observers for a “Survey on Judgment and Decision Making” (as well as the same number of allocators, which we describe in supplementary analyses on OSF). We eventually recruited 252 observers via MTurk. According to pre-registered exclusion criteria, we removed participants who failed to comprehend the multi-step instructions. Eleven observers fell below this criterion, resulting in a final sample of 241 observers ($M_{age} = 39.65$, 49.8% female). Importantly, as in Studies 1-2, the results remain consistent whether or not we enforce this pre-registered exclusion criterion.

WHEN WASTE PAYS

Design. Participants (hereafter “observers”) played a trust game in which they decided how much to entrust to another MTurker (hereafter “allocator”) who had previously chosen an equal allocation, and how much to entrust to an allocator who had previously chosen an efficient allocation. This “strategy method,” common in tests of interactive behavior, asked individuals to make their decisions for all possible allocator choices before the game occurred (Camerer, 2003; Jordan et. al., 2016). We randomized the order in which equal and efficient allocator choices appeared to observers. This resulted in a 2 (Allocator Choice: Equal, Efficient) X 2 (Order: Equal-first, Efficient-first) mixed-factorial design, with Allocator Choice as a within-subjects factor and Order a between-subjects factor.

Procedure. We first described the rules of the trust game, which we referred to only as a “game” in our instructions. We gave observers a \$.30 bonus to begin, then told them that they would have the chance to send any portion of that bonus, including none at all, to the allocator. We would triple what they sent, and allocators would then have the chance to return any portion of what they received. We quizzed participants on the rules of the game.

After initial game instructions, we described the equal or efficient allocation decision using the same language as in Study 2. The end of the allocation decision instructions stated that observers would be able to base how much they sent on the allocator’s allocation. As with the trust game instructions, we quizzed participants on their understanding of the allocation decision and its application to the game.

Dependent variable. Observers then decided how much money to transfer to an allocator making the efficient allocation and how much money to transfer to an allocator making the equal allocation (with order of decision varied by Order condition). Observers made their two trusting decisions on separate screens, using a sliding scale defaulted to zero. We hypothesized that

WHEN WASTE PAYS

observers would trust equal allocators more than efficient allocators, signifying the presence of a waste-pays effect.

Moderating variable. We then asked observers which allocation—equal or efficient—*they themselves* would have chosen and which allocation they thought would be easier to defend to others. As described above, we hypothesized that, while observers who prefer equality would draw negative inferences about allocators who disagree (i.e., demonstrating a waste-pays effect), observers who prefer efficiency would trust equal and efficient allocators the same.

To assess the stability of these preferences, we contacted observers approximately two months after they completed the initial survey and asked them again for their equal-versus-efficient preferences. Among these observers, 193 of 252 responded, and 78.2% of these 193 kept the same preference, suggesting reasonable stability of preferences.

Mediating variables. We measured perceptions of warmth and competence to assess whether these social perceptions mediated financial trust decisions. Observers completed twelve items from the warmth and competence subscales (Fiske, Cuddy, Glick, & Xu, 2002). We randomized the order of items and displayed four per screen; we phrased the items to compare an allocator who chose the equal allocation with an allocator who chose the efficient allocation. For example, we asked: “Who is more likely to be competent, a person who made the equal allocation or the efficient allocation?” The seven-point scale ranged from “*A person who chose equal would be much more competent* – *Neither would be more competent* – *A person who chose efficient would be much more competent.*” We re-scaled the scores to range from -3 to +3. We also re-scored the warmth and competence measures to reflect perceptions of an allocator making the same allocation decision as the observer (i.e., if an observer preferred the efficient allocation, we reversed the scoring of the warmth and competence measures). In sum, these

WHEN WASTE PAYS

variables represent warmth and competence perceptions *conditional on equal-versus-efficient allocation agreement*. More positive numbers indicate a stronger preference for agreeing others (i.e., allocators with the same preference as the observer), whereas more negative numbers indicate a stronger preference for disagreeing others.

Exploratory variables. To assess the effectiveness of the manipulation, we asked, via an open-ended, optional response, what the observers considered as they made their trust decisions and whether an allocator's choice of an equal or efficient allocation told the observers anything about the likelihood that an allocator would return money in the game. To assess the presence of any false consensus effect (i.e., the tendency to over-estimate the extent to which others agree with ones preferences; Ross, Greene, & House, 1977), we asked the participants to imagine 100 other mTurkers and then to imagine how many of them would choose the same option as that preferred by the observer. (We conditioned the wording on the observer's response earlier in the survey.) Finally, we collected a variety of other exploratory measures (results available on OSF), executed the trust game, and paid bonuses based on random pairings of observers and allocators.

Results

Analytic plan. According to our pre-registered analysis plan, we accounted for repeated measures in our data by using a multi-level regression with random intercepts for each subject. (Observers made trusting decisions for each equal-versus-efficient allocation the allocator could make.) Further, we assumed a beta distribution to approximate the error function in our regression equations. (See online materials for a comparison of normal and beta distributions overlaid on the distribution of observer trust decisions.) These analyses were conducted in R using the glmmADMB package (Fournier et al., 2012). We used this regression set-up for

WHEN WASTE PAYS

observer decisions only; for other measures in our data, we used standard linear regressions. Results remain consistent when we implement standard linear regressions for all analyses.

Manipulation checks. We found neither a main effect of Order (*log-odds 95% CI* [0.69, 1.20], $z = -0.65$; $p = .514$) nor an interaction between allocator Choice and Order (*log-odds 95% CI* [0.74, 2.22], $z = 0.92$; $p = .359$). Thus, we collapsed across Order for subsequent analyses.

Main effect of allocator choice. Our first hypothesis concerned the waste-pays hypothesis: Did observers infer that equal allocators were, on average, more trustworthy? To test this, we compared the amount awarded to allocators who favored equality with the amount entrusted to allocators who favored efficiency. Observers sent a higher proportion of their endowment to the former ($M = 59.3\%$) than the latter ($M = 50.9\%$; *log-odds 95% CI* [1.02, 1.64], *cohen's d* = .21, $z = 2.18$; $p = .029$). The effect of allocator choice on observer trust remained significant when we controlled for observer gender, age, religiosity, risk preferences, and own preference in the equal-versus-efficient allocation in a simultaneous regression (*log-odds 95% CI* [1.03, 1.65], $z = 2.21$; $p = .027$).

Moderation by observer preference. Our second hypothesis concerned whether observer inferences depended on their (the observer's) own preference. We predicted that, while observers who favored equality would punish allocators who disagreed (i.e., showing a waste-pays effect), observers who favored efficiency would not (they would trust equal and efficient allocators alike). To test this, we examined an interaction term between Allocator Choice and the observer's own preference for an equal or efficient allocation. This term was significant (*log-odds 95% CI* [1.28, 3.27], $z = 2.98$; $p = .003$).

To isolate the locus of the interaction, we compared the simple effects of allocator choice across different observer preferences. Notably, observers were approximately evenly split in

WHEN WASTE PAYS

their own preferences: 46% of observers preferred the equal allocation and 54% preferred the efficient allocation. Observers who preferred the efficient allocation sent statistically equivalent amounts to equal allocators ($M = 60.8\%$) and efficient allocators ($M = 64.5\%$; *log-odds 95% CI* [0.79, 1.54]; *cohen's d* = .09, $z = 0.59$; $p = .554$). However, observers who preferred the equal decision sent significantly more to equal allocators ($M = 58.1\%$) than they did to efficient allocators ($M = 39.5\%$; *log-odds 95% CI* [0.30, 0.98]; *cohen's d* = .51, $z = -2.02$; $p = .044$; See Figure 2).

For interested readers, we note that observers who preferred the *efficient* allocation were more trusting overall, transferring an average of 63% of their endowment as compared to an average of 49% for observers who preferred the equal allocation (*log-odds 95% CI* [1.03, 1.79], *cohen's d* = .34, $z = 2.20$, $p = .028$). We attribute this to the asymmetric punishment pattern of observers who preferred equality. They viewed allocators who disagreed as lacking trustworthiness, while efficient allocators viewed allocators who disagreed as equally trustworthy.

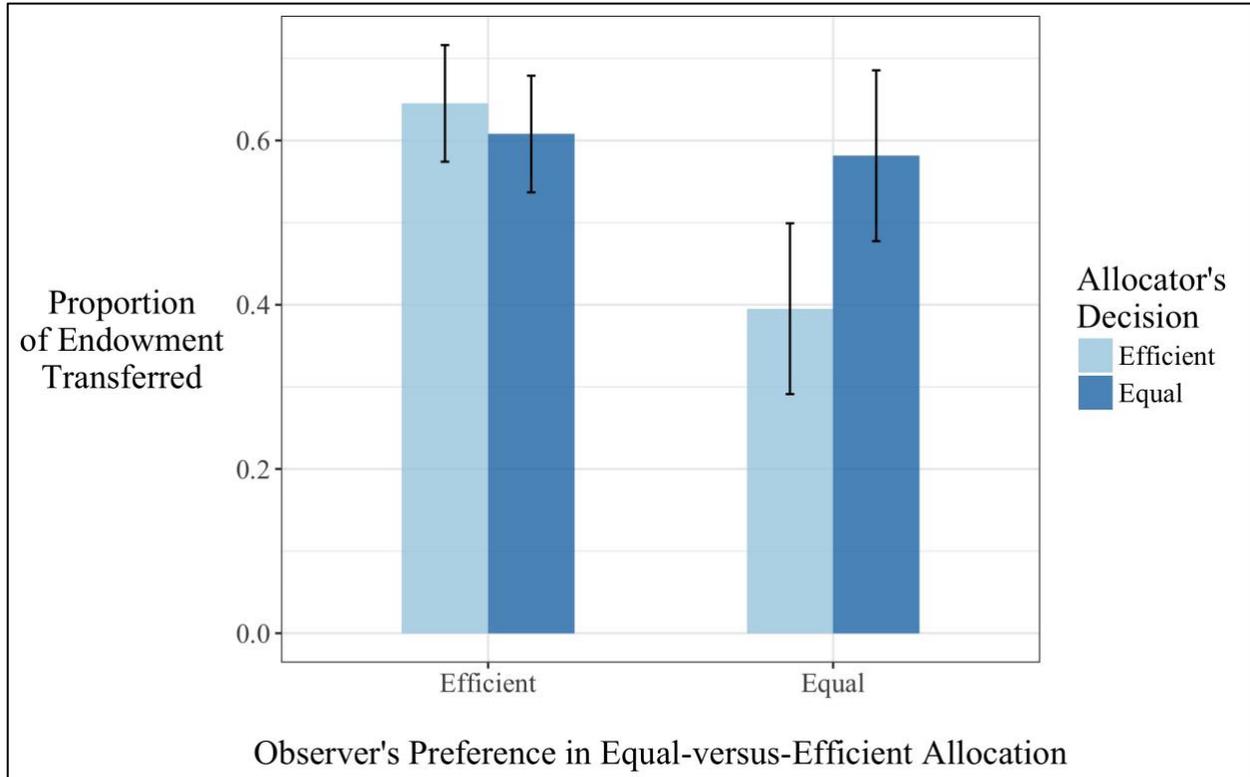


Fig. 2. The Y-axis displays the proportion of endowment transferred from observers in the trust game (Study 3). The left side of the panel shows the average amounts sent by observers who favored the efficient allocation, while the right side of the panel shows the average amounts sent by observers who favored the equal allocation. The two colors depict the allocator's possible allocation decisions. Light blue shows the amount transferred to allocators who chose the efficient allocation; dark blue shows the amount transferred to allocators who chose the equal allocation.

Social perceptions of warmth and competence. Our third hypothesis concerned whether asymmetric perceptions of warmth (but not competence) would statistically mediate asymmetric differences in trusting behavior. To analyze perceptions of warmth and competence, we averaged the six items from the warmth and competence subscales. As described above, these variables represent warmth and competence perceptions conditional on equal-versus-efficient allocation agreement, with more positive numbers indicating a stronger preference for an allocator who agreed with the observer.

Perceptions of warmth. An analysis of the composite warmth measure showed that observers who preferred the equal allocation believed that an allocator who made the equal allocation would be warmer than an allocator who made the efficient allocation (95% CI [1.20,

WHEN WASTE PAYS

1.61]; $t(130) = 13.44, p < .001$). Observers who preferred the efficient allocation did *not* believe that an allocator who made the efficient allocation would be warmer than an allocator who made the equal allocation; if anything, they believed the *opposite*, that those who disagreed on the appropriate allocation – preferring equality over efficiency – would be warmer (95% CI [-.49, -.04]; $t(109) = -2.31, p = .023$).

Perceptions of competence. Our measures of competence largely mirrored, in reverse fashion, these measures of warmth. Observers who preferred the *efficient* allocation perceived allocators who did the same as much more competent than allocators who made the equal allocation (95% CI [0.70, 1.10]; $t(109) = 9.05, p < .001$). Observers who preferred the equal allocation did not show the reverse preference; they perceived no difference in competence between the two types of allocators (statistically indistinguishable from zero: 95% CI [-0.09, .31]; $t(130) = 1.11, p = .268$).

Mediation by social perceptions of warmth and competence. To analyze the relationships among warmth, competence, and trust, we defined a new outcome measure: the difference between how much an observer transferred to an allocator with the *same* equal-versus-efficient preference as herself, as compared with how much an observer transferred to an allocator with a *different* equal-versus-efficient preference as herself. See Figure 3 for a diagram of the complete model.

Path 1: Do perceptions of warmth mediate trust? In line with our pre-registered analysis plan, we conducted a between-subjects mediation analysis using a parallel multiple mediator model to simultaneously assess the mechanisms of warmth and competence. We found evidence for the hypothesized indirect effect of warmth. First, observers who preferred the equal allocation demonstrated a much stronger effect of agreement on perceptions of warmth than did

WHEN WASTE PAYS

observers who preferred the efficient allocation (95% CI [1.46, 1.89], $z = 15.31$; $p < .001$). In turn, this difference in perceptions of warmth was associated with trusting behavior (95% CI [1.14, 2.49], $z = 5.23$; $p < .001$). Taken together, we found strong evidence for an indirect path: 95% CI [1.84, 4.24], $z = 4.95$; $p < .001$.

Path 2: Do perceptions of competence mediate trust? We also found evidence for an opposing indirect path through competence. Specifically, observers who preferred the efficient allocation demonstrated a stronger effect of agreement on perceptions of competence than did observers who preferred the equal allocation (95% CI [-0.99, -0.59], $z = -7.78$; $p < .001$). In turn, this difference in perceptions of competence was associated with trusting behavior (95% CI [0.34, 1.81], $z = 2.88$; $p = .004$). Taken together, we found evidence for an indirect path: 95% CI [-1.46, -0.23], $z = -2.70$; $p = .007$. While we cannot make causal claims using these associative measures, we observed associations consistent with the possibility of *oppositional mediation*, in which two indirect paths exert opposing influences on the dependent variable of interest.

Paths 1 and 2: Overall test of the mediation model. While both indirect paths reached traditional levels of statistical significance, a pairwise contrast between the two indirect effects revealed that the indirect path through warmth was approximately *triple* the size of the indirect path through competence ($b = 3.04$ vs. $b = -0.85$), and this difference was significantly different from zero; contrast between the two paths: 95% CI [0.65, 3.74], $z = 2.78$; $p = .005$). Thus, while the analyses are consistent with the conclusion of *oppositional mediation*, perceptions of warmth carried far more explanatory power than did perceptions of competence in predicting the asymmetric effect of agreement on trust behavior.

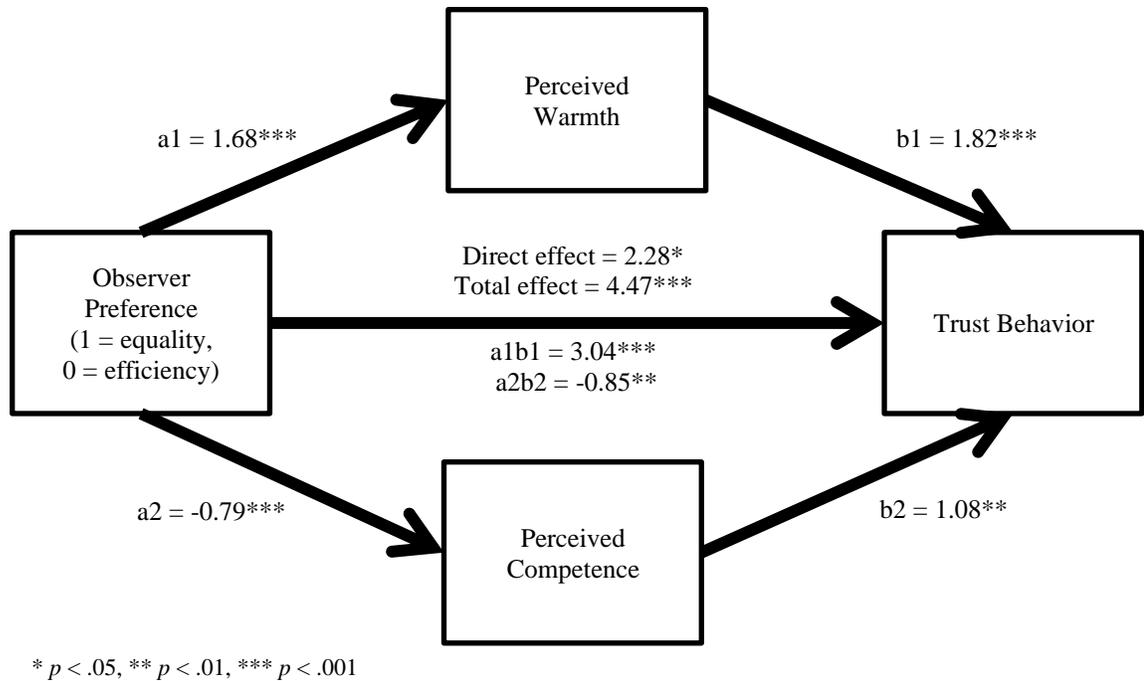


Fig. 3. Path analyses of the role of warmth and competence in trusting behavior in Study 3. We observed an oppositional mediation pattern, though perceptions of warmth carried far more explanatory power than did perceptions of competence in predicting the asymmetric effect of agreement on trust behavior.

Discussion

On average, observers drew harsher inferences about efficient allocators than they did about equal allocators, indicating the presence of a *waste-pays effect*. Crucially, however, observers' inferences were qualified by their own views. Observers who preferred the efficient allocation trusted both types of allocators to the same degree. By contrast, observers who preferred the equal allocation viewed *only* a preference for equality as a signal of an allocator's social warmth and, in turn, trusted only those allocators who chose equality.

General Discussion

Three pre-registered experiments (N=1,095) extended the revised value pluralism model (Tetlock et al., 1996) by testing the effect of social evaluation on equality-efficiency trade-offs. Concerning allocator behavior, results supported the *wasteful-when-watched hypothesis*: social evaluation increased equal (yet inefficient) decisions involving hypothetical organs (Study 1) and

WHEN WASTE PAYS

real money, despite the fact that an unequal allocation would have been Pareto-optimal (Study 2). Concerning observer behavior, results supported the *waste-pays hypothesis*: observers financially rewarded allocators who favored equality over efficiency (Study 3).

These findings extend the RVPM, revealing that social evaluation from observers prioritizes equality relative to efficiency. While prioritizing equality may often be socially desirable, in the circumstances studied here, social evaluation triggered undesirable outcomes. Organs were not given to those most able to benefit from them and money was not used to create maximal value, even though to do so would have increased overall value and made no one worse off.

Importantly, the idea that social-structural factors dynamically shape tradeoff reasoning suggests qualifications to such philosophical perspectives as virtue ethics, which typically characterizes moral values as stable dispositions across time and contexts (Hursthouse 2016). In contrast to this view, and consistent with pioneering views in psychology (e.g., Choshen-Hillel, Shaw, & Caruso, 2015; Halevy & Chou, 2014), we hypothesized and observed not only that value hierarchies respond to momentary social evaluation but also that they do so even when real stakes were on the line. More practically, the present results reveal an important caution for the design of allocation systems: Social factors can hinder the chances that scarce resources will be allocated in an efficient way, even when the efficient way creates greater value than at the outset and leaves no one worse off.

Limitations. While cross-cultural variation in value hierarchies is a rich area of research (cf. Inglehart, 1977), the present work contains only participants from the United States. It may be that in countries where a value of avoiding waste (efficiency) more strongly trumps the value of equality, thresholds for switching to equality may be lower (for related work, see Awad et al.,

WHEN WASTE PAYS

2018). A second limitation concerns types of social evaluation. Existing research makes clear that different types of accountability influence judgment and choice in different ways (Lerner & Tetlock, 1999). In the present research, we focused on social evaluation by third-party observers. An important area for future research is to investigate how different types of accountability alter tradeoff reasoning, in general, and equality-efficiency tradeoffs, in particular.

Summary. The present research sheds new light on the effect of social evaluation on value trade-offs. Results reveal that social observation leads individuals to make inefficient (yet equal) allocations regardless of whether the scarce resource is human organs or money. Moreover, such inefficient (yet equal) allocations confer social and economic benefits. Taken together, these experiments elucidate multiple pathways through which the social context shapes trade-offs between colliding values, in this case creating undesirable outcomes.

References

- Adelberg, S., & Batson, C. D. (1978). Accountability and helping: When needs exceed resources. *Journal of Personality and Social Psychology*, *36*(4), 343–350.
<https://doi.org/10.1037/0022-3514.36.4.343>
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59.
- Becker, G. (1981). Altruism in the Family and Selfishness in the Market Place. *Economica*, *48*(189), new series, 1-15. doi:10.2307/2552939
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. <https://doi.org/10.1006/game.1995.1027>
- Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2015). Waste management: How reducing partiality can promote efficient resource allocation. *Journal of Personality and Social Psychology*, *109*(2), 210.
- Everett, J. A. C., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, *79*, 200–216.
<https://doi.org/10.1016/j.jesp.2018.07.004>

WHEN WASTE PAYS

- Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*(6), 772–787. <https://doi.org/10.1037/xge0000165>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. <https://doi.org/10.1037//0022-3514.82.6.878>
- Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, *111*(Supplement 4), 13593–13597. <https://doi.org/10.1073/pnas.1317505111>
- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, *18*(2), 255–297. <https://doi.org/10.1111/0162-895X.00058>
- Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*(2), 233–249. doi: 10.1080/10556788.2011.597854
- Gordon-Hecker, T., Choshen-Hillel, S., Shalvi, S., & Bereby-Meyer, Y. (2017). Resource allocation decisions: When do we sacrifice efficiency in the name of equity? In *Interdisciplinary Perspectives on Fairness, Equity, and Justice* (pp. 93–105). Springer, Cham. https://doi.org/10.1007/978-3-319-58993-0_6

WHEN WASTE PAYS

- Halevy, N., & Chou, E. Y. (2014). How decisions happen: Focal points and blind spots in interdependent decision making. *Journal of Personality and Social Psychology*, *106*(3), 398–417. <https://doi.org/10.1037/a0035351>
- Hochman, H., & Rodgers, J. (1969). Pareto Optimal Redistribution. *The American Economic Review*, *59*(4), 542-557. Retrieved from <http://www.jstor.org/stable/1813216>
- Hursthouse, R. (2016). Virtue Ethics. *Stanford Encyclopedia of Online Philosophy*. <https://plato.stanford.edu/entries/ethics-virtue/>
- Inglehart, R. (1977). Long Term Trends in Mass Support for European Unification. *Government and Opposition*, *12*(2), 150–177. doi: 10.1111/j.1477-7053.1977.tb00529.x
- Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, *65*(1), 19-51.
- Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*, *113*(31), 8658–8663. <https://doi.org/10.1073/pnas.1601280113>
- Kant, I. (1797/1948). *Groundwork on the Metaphysics of Morals*. Cambridge Texts in the History of Philosophy, edited by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- LeVeck, B. L., Hughes, D. A., Fowler, J. H., Hafner-Burton, E., & Victor, D. G. (2014). The role of self-interest in elite bargaining. *Proceedings of the National Academy of Sciences*, *111*(52), 18536-18541.

WHEN WASTE PAYS

- Lichtenstein, S., Gregory, R., & Irwin, J. (2007). What's bad is easy: Taboo values, affect, and cognition. *Judgment and Decision Making*, 2(3), 169-188.
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279–301. doi: 10.1016/0022-1031(77)90049-x
- Rosseel, I. (2011). Lavaan: an R package for structural equation modeling and more Version 0.4-9 (BETA). *Journal of Statistical Software*.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404.
<https://doi.org/10.5465/amr.1998.926617>
- Sacco, D. F., Brown, M., Lustgraaf, C. J. N., & Hugenberg, K. (2017). The adaptive utility of deontology: Deontological moral decision-making fosters perceptions of trust and likeability. *Evolutionary Psychological Science*, 3(2), 125–132.
<https://doi.org/10.1007/s40806-016-0080-6>
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology* (Vol. 25, pp. 1-65). Academic Press.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>

WHEN WASTE PAYS

- Suedfeld, P., Bluck, S., Loewen, L., Elkins, D. (1994). Sociopolitical values and integrative complexity of members of student political groups. *Canadian Journal of Political Science*, 26, 121-141.
- Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology*, 50(4), 819–827. <https://doi.org/10.1037/0022-3514.50.4.819>
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of personality and social psychology*, 78(5), 853.
- Tetlock, P.E., Micheletti, P., & Hannum, K. (1984). Stability and change in senatorial debate: Testing the cognitive versus rhetorical style hypothesis. *Journal of Personality and Social Psychology*, 46, 979-990.
- Tetlock, P. E., Peterson, R. S., & Lerner, J. S. (1996). Revising the value pluralism model: Incorporating social content and context postulates.
- Ubel, P. A., & Loewenstein, G. (1996). Distributing scarce livers: The moral reasoning of the general public. *Social Science & Medicine*, 42(7), 1049–1055. [https://doi.org/10.1016/0277-9536\(95\)00216-2](https://doi.org/10.1016/0277-9536(95)00216-2)
- Uhlmann, E. L., Zhu, L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, 126(2), 326–334. <https://doi.org/10.1016/j.cognition.2012.10.005>