The Role of Meta-Context in Moral Decisions[*]

Folco Panizza[†] **Alexander Vostroknutov**[‡ §] **Giorgio Coricelli**[¶]

January 18, 2021

Abstract

Members of discriminated groups often make moral decisions based solely on the context of allocations that they believe are feasible for them. Such contexts can be seriously constrained by various physical, economic, or institutional factors—e.g., no access to education, high costs of education, or community norms that proscribe getting it—which can lead to normative acceptance of discriminated societal roles. In a lab experiment, we study how easily moral judgements can be affected by the *meta-context*, or allocations that are not present in the current context, but were experienced by the decision maker in the past (e.g., an experience of an unconstrained educational choice). We find that moral judgements are strongly affected by such experiences. This mechanism suggests that making discriminated groups realize that they are discriminated can create a powerful normative incentive for them to change their situation.

JEL classifications: *C91, C92, D91.*
Keywords: *social norms, meta-context, discrimination, moral judgements.*

---

[*]We would like to thank the participants of the Florence-Konstanz Workshop on Behavioral Social Sciences (Nov 2017) and seminars in Vienna Center for Experimental Economics (Oct 2017) and Max Planck Institute for Research on Collective Goods (May 2017) for invaluable comments. GC acknowledges the financial support of the European Research Council (ERC Consolidator Grant 617629). All mistakes are our own.

[†]Center for Mind/Brain Sciences, University of Trento, Via delle Regole 101, 38123 Mattarello (TN), Italy. e-mail: folco.panizza@unitn.it

[‡]Department of Economics (MPE), Maastricht University, Tongersestraat 53, 6211LM Maastricht, The Netherlands. e-mail: a.vostroknutov@maastrichtuniversity.nl

[§]Corresponding author.

[¶]Department of Economics, University of Southern California, 3620 South Vermont Ave. 300, Los Angeles, CA 90089, USA.

1 Introduction

"If they had seen what we see, they would have judged as we judge." – Galileo Galilei

In a recent interview to Financial Times (Greely, 2020) IMF chief Kristalina Georgieva remembers how back in the 80's in then communist Bulgaria she helped receiving a Japanese delegation. After what she thought was a lavish reception, one of the Japanese told her: "You are lucky you don't know how poor you are." Georgieva mentions that this revelation made her think about the roots of poverty for the first time, which eventually shaped her career path. Back then, she came to the conclusion that the persistence of poverty in Bulgaria was not due to scarcity of resources, but rather low productivity and non-existence of competition. However, another important conclusion can be made from the observation of the Japanese delegation member. He pointed out that, in the countries under Soviet umbrella, the absence of knowledge about the outside world had led to the acceptance of poor economic conditions. Continuing this line of reasoning, we can entertain a possibility that the persistence of poverty—or at least the absence of discontent about it—may have had another source than low productivity. Locked in the informational universe promulgated by the Soviet propaganda machine, people did not realize that they lived in (relative) poverty, and consequently did not take any steps to change their lives (e.g., choosing a more democratic government). After the fall of Soviet regimes in the 90's, George Soros' Open Society Institute and other similar organizations actively promoted "civil society" in the post-Soviet space following similar intuition (Henderson, 2003).

Situations like this—characterized by the moral acceptance of one's adverse circumstances due to the belief that better alternatives are infeasible—are not a thing of the past. They might still be a significant factor that contributes to the persistence of poverty, discrimination, and lack of inclusivity in many areas of economic activity. For example, even though young girls today can see many successful female scientists on the news and various TV shows, they might still have *low aspirations* to scientific career, especially if they come from a community where it is believed that girls are supposed to spend most of their time caring for their families instead of studying or if there are physical constraints like limited access to good education (Correll, 2001; Akos et al., 2007).[1] As a result, they might accept the family-oriented scope of their lives and do not even think about putting effort into pursuing a scientific career (e.g., Isik et al., 2018). In such circumstances, an economic policy that aims at facilitating female participation in science or other professions might not be particularly effective because the target population does not consider such a career as a possibility, and moreover might be against it on the moral grounds that it undermines the "traditional" female occupations (Steuter, 1992). Thus, discrimination can adversely affect vulnerable social groups not only through conventionally considered channels that involve instances of discriminatory behavior (Rodgers, 2009; Becker, 2010; Lane, 2016),

[1] Another reason for not considering certain alternatives as possible might be narrow bracketing (Read et al., 1999) that makes people think in terms of proximate instead of distal goals.

but also through the indirect effect on normative beliefs that disincentivize the members of discriminated groups from participating (Ceci and Williams, 2011; Beaman et al., 2012; Breda et al., 2020).

The question we tackle in this study is how hard it is to *change* normative beliefs in situations like these. Can we convince people in a developing country that their governments keep them in informational darkness and that their economic conditions are morally unacceptable when seen from the perspective of an open society? Can we convince young girls that they should try to become scientists even when it seems like an impossible feat? Paraphrasing Galileo, can we help them see what we see, so that they can judge as we judge? Following a recent theory of social norms (Kimbrough and Vostroknutov, 2020, further KV), we conjecture that humans already have a "built-in" capacity to base their moral judgements on a *wider scope of possibilities than those they consider available to them*, the meta-context. People in developing countries, given enough information about or experience of such wider meta-context, can get dissatisfied with how their governments treat them and take steps to change the situation (see Faris, 2013, for the case of Arab Spring). Young girls can become aware that local community norms restrict their access to education, find it inappropriate, and, as a result, put more effort into pursuing their dreams (Beaman et al., 2012). All this being said, it is not the goal of this study to estimate the general importance of factors like these in the success or failure of inclusivity or anti-discrimination policies. Rather, as the first step and the proof-of-the-concept, we propose and experimentally test a theoretical framework that allows to quantify the effects of meta-context and to determine whether they influence behavior in principle. Our goal is therefore to test the theory that predicts meta-context effects.

To understand our intuition, consider a decision maker who chooses a division of a fixed amount of some resource between herself and other players in the *context* of some Dictator-like game. By "context" we mean the set of all allocations that are achievable in it. For example, it can be a choice of how to divide a fixed amount of time (e.g., working hours) between studying and helping your family. In "open" societies, the context of this choice will consist of all possible divisions of time: the decision maker is free to dedicate all her time to studying, all her time to caring for the family, or any combination in between. In such circumstances, the decision maker, who follows social norms, will choose to spend, say, half of her time studying and the other half helping her family (giving half of the resource to the other player, as is often observed in experimental Dictator games, Engel, 2011). However, in more "traditional" societies it might be considered unthinkable for a woman to spend less than half of her time helping her family (see e.g., Voicu et al., 2009). Here, the context of the decision is different: it is possible to allocate to studying only up to a half of your time. The allocations where more than half of the time is dedicated to studying are not available anymore (at least in the mind of the decision maker). In this case, the same norm-following individual might optimally allocate only one quarter of her time to studying. Notice that in both cases the decision maker chooses the most socially appropriate allocation and finds nothing morally wrong with her decision: this is how things are supposed to be.

To put this argument into mathematical terms, we use the theory of injunctive norms proposed by KV. It suggests that pro-social behavior results from the general propensity to follow injunctive norms that are determined by the context of a game being played (context in the sense mentioned above). It is assumed that player i maximizes norm-dependent utility $u_i(x) + \phi_i\eta(x)$, where $u_i(x)$ is the consumption utility in allocation x, $\eta(x)$ is the norm function that determines the normative valence of x in the interval $[-1, 1]$, and $\phi_i \geq 0$ is i's general propensity to follow norms.[2] According to this utility specification, there is a trade-off between consumption utility (from studying) and following the norms (helping your family).

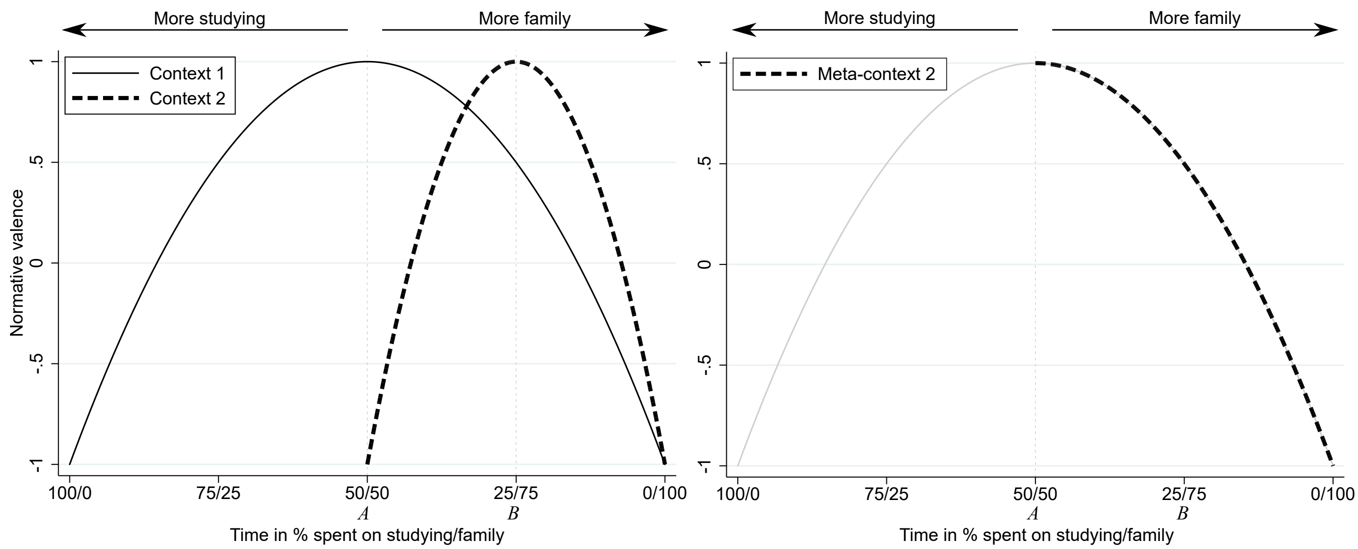

Figure 1: **Left panel.** The norm functions in the two contexts. **Right panel.** The meta-context norm function in context 2.

The left panel of Figure 1 shows two norm functions arising from this theory in the two contexts mentioned above. In context 1, the decision maker can choose any allocation from 100/0 (all time to studying) to 0/100 (all time to family). The solid-line parabola represents the normative valences of each allocation in context 1, measured from -1 (very inappropriate) to 1 (very appropriate). A rule-following decision maker with high enough ϕ_i will choose allocation A (50% of time to studying) that corresponds to the highest level of social appropriateness. In context 2, only the allocations that give at least 50% of the time to family are available (or believed to be available). In this context, the dashed-line parabola describes the normative valences according to the theory. In this case, a rule-following decision maker chooses allocation B (25% of time to studying) that is the most socially appropriate in this context. It is this decision that creates a problem. In context 2, the rule-following decision maker believes that B is the best choice that

[2]The norm function η is calculated using the information on all utilities that all players get in all allocations. See Section 3 below.

maximizes her norm-dependent utility. Moreover, according to this, any attempts to increase the amount of studying without changing the perceived context will be considered as the attempts to make the decision maker do something that goes against her moral views, namely to choose allocations to the left of B that do not have the highest social appropriateness in context 2. So, if an economist decides to introduce a policy that simply decreases the costs of education for women in context 2, such a policy will probably not increase the time dedicated to studying all the way up to 50%, as in the open society represented by context 1. The reason is that this level of studying is considered extremely inappropriate, so we would expect that the policy might not reach its goal of pushing studying time in context 2 to the level of context 1.

While this argument is straightforward and simply suggests that certain cultural or physical constraints might influence educational choices in some specific ways, it remains unclear what happens when women living in context 2 learn about the existence of context 1. Specifically, when they are presented with the *meta-context* that includes their own context 2 and some other allocations unavailable in it. Suppose that they learn that in the neighboring country women can choose to dedicate any amount of time to education instead of only half (context 1). In the optimistic scenario, the situation will change as shown on the right panel of Figure 1. Here women perceive the normative valences of allocations in context 2 as if coming from context 1 that, according to their new belief, represents the set of possibilities that they should have. Notice that now point A becomes the most socially appropriate action. This change in beliefs will make women think that the treatment they receive in their own society (context 2) is unfair and should be changed. In the pessimistic scenario, they will not react to this information at all since, in the end, the allocations enjoyed by someone else in context 1 are not "available" to them.[3]

Knowing how much meta-context influences social decisions like those above is important. People, who observe or experience meta-context allocations that are typically infeasible for them, might change their normative viewpoint on their current situation, which can lead to dissatisfaction and consequently—at least in democratic societies—to a change in policy that accommodates the updated normative position of the population. Thus, the ability to take meta-context into account can lead to desirable societal changes. However, in order to understand if this route has a potential for successful policy-making, we need to know the extent to which meta-context can influence social decisions.

[3]Some evidence from the lab suggests the optimistic scenario. In the experiment of List (2007), subjects are given \$5 to keep and another \$5 that they can share with someone else. The modal amount given is \$2.5. In Thomsson and Vostroknutov (2017), subjects are given \$10 and are told that they can share it with another subject, but with the restriction that no more than \$5 can be given. In this case, the modal amount given is \$5, even though the two games are identical from the game-theoretic perspective. The difference in behavior may come from the different perception of the meta-context: while in the former experiment subjects are told that only \$5 is for sharing (like context 2 in the left panel of Figure 1), in the latter experiment the subjects are told that \$10 is for sharing, but there is a restriction on the amount that can be given (like meta-context 2 in the right panel of Figure 1).

The goal of our study is therefore twofold. First, we want to determine whether or not meta-context has any effect on social decision making at all. Second, if it does, we would like to know if it is "easy" or "hard" for people to switch their normative reasoning from the current context to the meta-context. The answer to the first question will establish whether the argument, we have presented so far, has any merit. The answer to the second question is crucial for understanding which policies should be used to promote certain behaviors in the situations where context is restricted. If it is easy to induce thinking in terms of the meta-context, as we hypothesize, then all we need to do is to convince the targeted group of people that there are alternatives that they can and should consider.

To answer these questions we have designed an experiment that is not meant to imitate any real-life choices, but rather constitutes the simplest and the most uncontroversial test of the meta-context effect described above as well as the theory of KV on which it is based. In the experiment, each participant is presented with 182 randomly-ordered choices between two allocations that assign certain amount of money to her and some other player. The sum of payoffs in all allocations and all periods is constant and equal to 60 points (later exchanged for money). For example, in some period t a participant might be presented with a choice between $(40, 20)$ and $(10, 50)$, two allocations with the first payoff going to the participant, and the second to another randomly chosen player. The idea behind this design is that, in each period t, the decision maker chooses only between two allocations, however, all the allocations that she has experienced in the past, namely those in periods from 1 to $t - 1$, together constitute the meta-context of the current decision and can exert some detectable influence on it. To test for this possibility, we use the theory of KV to compute the individual norm functions in each period t that are based on all allocations that the decision maker has experienced up to that point and then plug the resulting independent variable into a logit regression of the observed binary choices. Then we compare how this model fares in comparison to other specifications.

We find that the meta-context consisting of the allocations experienced in the past does in fact exert influence on the decisions of our participants in the way predicted by the theory: participants do take meta-context into account when making moral judgements. The meta-context regression model outperforms 18 other specifications that we check. This strongly suggests that simply experiencing a choice between certain allocations in the past includes these allocations into the moral reasoning related to the future social decisions. We believe that this process is mostly "automatic" and probably similar to regret avoidance (see Fioretti et al., 2021, and discussion). This is a good news for policy makers who deal with problems related to openness and discrimination as those showcased above. Our results suggest that *experiencing* allocations, or having a choice among them, changes moral reasoning even though these experienced allocations are not available anymore. This points to a specific type of policies that can help people to see what they can achieve and to create intrinsic normative incentives to eliminate inequality and discrimination.

Overall, our experiment corroborates the general theory of KV, which means that it can be used to analyze the effects of meta-context in various game-theoretic settings. For example, this framework can be applied to predicting the influence of anti-discrimination information campaigns on the participation of the target population in certain activities (e.g., education) or to estimate the effects of online presence that makes information about the world available to populations in developing countries. We believe that better understanding of normative beliefs and the ways to change them can be indispensable for the new type of economic policies that are designed to create normative incentives instead of monetary ones.

2 Experimental Design

The experiment consisted of three decision-making stages: the Rule Following (RF) task, the series of mini-Dictator games, and the norm-elicitation task (instructions can be found in Appendix E).

2.1 The Rule-Following Task

In the rule-following task (Kimbrough and Vostroknutov, 2018), participants decided how to allocate 50 balls between a blue and a yellow bucket. The position of the two buckets was randomised across individuals. Participants earned €0.05 for each ball they dropped in the blue bucket, and €0.10 for each ball they dropped in the yellow bucket. The instructions explicitly stated that *'the rule is to put the balls into the blue bucket.'* This payment scheme and the rule were the only information provided to participants. The total payoff in this stage was the sum of earnings from both buckets. Therefore, the amount of money earned could vary from €2.50 (always following the rule) to €5.00 (never following the rule).

The RF task creates a situation in which participants are asked to follow an arbitrary rule that decreases their payoff and yet entails no cost of breaking it. This allows us to measure the propensity of participants to stick to a rule that, as was demonstrated by Kimbrough and Vostroknutov (2016), also predicts pro-social behaviour.

2.2 The Mini-Dictator Games

In Stage 2, participants played a series of mini-Dictator games with constant sum of payoffs. In each game each participant chose between two divisions of 60 tokens (1 token = €0.10) split between her and an unknown other. 27 different allocations were used in all mini-DGs, including an equal split. In 13 allocations the dictator received more than the recipient, and in other 13 allocations the dictator received less than the recipient. The allocations were combined to form a mini-DG according to the two criteria: 1) at least one of the two allocations has to be equal split or give to the dictator more than to the recipient; 2) if in one allocation the dictator receives less money than the recipient, then this allocation has to be less unequal than the other allocation (in which the dictator always receives more than the recipient). These criteria were chosen to make the selfishness/prosociality trade-off more salient and have yielded a total of 182 mini-DGs. The list of all mini-DGs can be found in Table 6, Appendix D. Mini-DGs were presented to each participant in an individually generated random order.

After the task was completed, participants were randomly paired and one of them was selected as a dictator. Consequently, one of the choices of this individual was randomly implemented. Participants were fully informed about all these procedures.

2.3 Norm Elicitation

In order to elicit participants' beliefs about norms we used the norm-elicitation task proposed by Krupka and Weber (2013), further the KW task. Participants were presented with a selection of 18 mini-DGs that they encountered in the previous task. They rated on a 4-point Likert scale the degree of social appropriateness of picking each alternative in each game. Each mini-DG was presented on a separate screen, and the order of their appearance was randomised. To detect choice-set effects, half of the allocations were repeatedly presented in combination with different alternatives. The list of all mini-DGs used for norm elicitation can be found in Table 7 in Appendix D.

To incentivize precise answers, participants were rewarded by means of a coordination game. One option in one of the mini-DGs was randomly selected. If participant's rating matched that of the majority in the session, he/she received €5.00.

2.4 Subjects

Participants were recruited from the subject pool of the Cognitive and Experimental Economics Laboratory at the University of Trento and invited via e-mail. 166 subjects (93 female, mean age = 22) completed the experimental task. The study was approved by the University Ethical Committee. Experiment was programmed in z-Tree (Fischbacher, 2007). Experimental sessions were run in May 2017, February and March 2018. There were no pilots, and no data were discarded.

3 Empirical and Theoretical Framework

In this section we present the empirical and theoretical framework that ultimately allows us to estimate the effects of meta-context on behavior. Consider a choice between two allocations $D = (d, 60 - d)$ and $A = (a, 60 - a)$ that represent two divisions of the pie of size 60 (as in the experiment). Here, the first material payoffs, d and a, go to the decision maker (the "mini-dictator") and the rest, $60 - d$ or $60 - a$, to the recipient. Assume without loss of generality that $d < a$. So, D is the *disadvantageous* allocation for the decision maker and A is the *advantageous* one. Suppose that the decision maker believes that both she and the recipient derive some strictly monotonic consumption utility $u(x)$ from a material payoff x. We consider the norm-dependent utility function (Kessler and Leider, 2012) of the dictator i:

$$U_i(D) = u(d) + \phi_i \eta(D) \quad \text{and} \quad U_i(A) = u(a) + \phi_i \eta(A). \tag{1}$$

Here ϕ_i is i's propensity to follow norms (Kimbrough and Vostroknutov, 2016) and η is a norm function that will be determined in different ways depending on the set-up.

Suppose we observe many choices of many subjects in different mini-DGs in different periods. Denote by D_{it} and A_{it} the two allocations between which subject i is choosing in period t. Then, given some fixed functional form of u and some definition of η for all allocations, we can estimate a random-effects logit regression of the form

$$\Pr[\text{choose } D_{it}] = \Lambda[c + \beta(u(d_{it}) - u(a_{it})) + \phi(\eta(D_{it}) - \eta(A_{it}))]. \tag{2}$$

Here $\Pr[\text{choose } D_{it}]$ is the probability of choosing the disadvantageous allocation that gives the dictator less payoff, Λ is the logistic cdf, and c, β, ϕ are the estimated coefficients. Following the random utility model of McFadden (1976), we assume that the probability of choosing the disadvantageous allocation is proportional to the norm-dependent utility difference between D_{it} and A_{it}, which boils down to estimating the regression coefficients on three independent variables: a constant, $u(d_{it}) - u(a_{it})$, and $\eta(D_{it}) - \eta(A_{it})$. The last two are the functions of the material payoffs or the beliefs elicited in the KW task.

The first regression that we report below in Section 4 estimates the regression model (2) using the consumption utility $u(x) = x^\alpha$ and the individual norm functions η_i obtained by interpolating the beliefs of subject i elicited in the KW task (see Appendix A).[4] In the next step, we estimate several regressions in which the definition of η comes from the theory instead of the elicited beliefs. Specifically, the model of KV constructs the normative valences attached to allocations D_{it} and A_{it} by using the information about the material payoffs implied by D_{it} and A_{it}, and possibly the material payoffs observed in the past.

The model relates the normative valences $\eta(D)$ and $\eta(A)$ to the *dissatisfactions* that both players feel when one allocation is chosen. The dissatisfaction that a dictator feels if the disadvantageous allocation D is chosen over A is equal to $u(a) - u(d) > 0$, the difference in the utilities obtained from her material payoffs. The recipient does not feel any dissatisfaction if D is chosen because it gives him the highest material payoff possible in the context of D and A. However, the recipient experiences dissatisfaction $u(60 - d) - u(60 - a) > 0$ when option A is chosen,

[4]The reason for estimating non-linear consumption utility will become clear below.

which gives him the materially worst payoff in the context of D and A (the dictator does not feel dissatisfaction when A is chosen). According to KV, the normative valences $\eta(D)$ and $\eta(A)$ derived from these dissatisfactions are proportional to the negatives of the sums of dissatisfactions of both players and can be computed as

$$\eta_c(D) = -\frac{1}{2}(u(a) - u(d)) \quad \text{and} \quad \eta_c(A) = -\frac{1}{2}(u(60-d) - u(60-a)).$$

Here the subindex c demarcates the *context model*, which takes into account only the *current context* of the decision consisting of allocations D and A. We also divide the negative dissatisfactions by 2, the number of allocations available, in order to be consistent with the later derivation of meta-context norm function that takes into account any arbitrary number of previous allocations.[5]

In order to use $\eta(D)$ and $\eta(A)$ to analyze our data, we need to make some assumptions about the utility function u. As KV have pointed out, in the case of two allocations, if u is linear then $\eta(D) = \eta(A)$, which makes no predictions about the choices. Thus, in order to test if subjects choose according to the context model we need to assume some concavity of u. We consider a class of utility functions $u(x;\alpha) = x^\alpha$ with $\alpha \in (0,1)$. In this case, the condition $\eta_c(D) > \eta_c(A)$, which states that the disadvantageous allocation is more appropriate than the advantageous one, can be rewritten as

$$d^\alpha + (60-d)^\alpha > a^\alpha + (60-a)^\alpha.$$

Notice that $x^\alpha + (60-x)^\alpha$ is a symmetric function with the maximum at $x = 30$ regardless of the value of α, akin to a parabola with the "horns down." Thus, the inequality above is satisfied whenever d is closer to the equal split than a (for all α). More generally, the prediction of this model—in the context of D and A—is that the more equal allocation is always more appropriate (as was suggested by Fehr and Schmidt (1999) and Fehr and Schurtenberger (2018)).[6] Thus, if we believe that the subjects in our experiment decide taking into account the *current context* of each separate decision (two allocations), we should be able to detect this by using $\eta_c(D_{it};\alpha) - \eta_c(A_{it};\alpha)$ as an independent variable in the regression model (2).

To give an example of the computation of η_c, consider the allocations $D_t = (30,30)$ and $A_t = (50,10)$ shown in Figure 2. The solid line connects the normative valences $\eta_c(D_t;0.9)$ and $\eta_c(A_t;0.9)$ normalized so that the lowest valence is zero (adding or subtracting a constant from η_c does not change anything in the estimation since only the differences of η_c enter the regression). We see on the graph that the more equal allocation $(30,30)$ has higher normative

[5]This is a simplified version of the re-normalization procedure in KV necessary to compare normative valences in different contexts with varying numbers of consequences.

[6]It can also be seen as a preference for maximin (Engelmann and Strobel, 2004), since a more equal allocation, in the specific context of dividing a pie, always produces a higher minimal payoff. In general, KV's model produces different "social preferences" for different contexts, so it is not helpful to generalize these seemingly "inequity averse" preferences to other situations.

valence than $(50, 10)$. So, the more equal allocation is more appropriate than the less equal one, as predicted analytically.

Notice that since the context model only takes into account the two allocations in period t, this model's predictions do not depend on *when* the choice is made. Specifically, the difference in normative valences $\eta_c(D_t; \alpha) - \eta_c(A_t; \alpha)$ only depends on the material payoffs in the current period and nothing else. It does not depend, for example, on whether the choice is made in period 1 or in period 182. However, as we claimed in the Introduction, there is a reason to expect that subjects might take *meta-context* into account, which consists of the allocations experienced in the past.[7] Consider any set of allocations $L = \{(\ell_k, 60 - \ell_k) \mid k = 1..K\}$ that includes 1) D and A, the two allocations between which the decision maker can actually choose, and 2) other allocations that were available before. Following KV, in this meta-context we can calculate the normative valences of D and A as

$$\eta_L(D) = -\frac{1}{K}\sum_{k=1}^{K} \max\{u(\ell_k) - u(d), 0\} + \max\{u(60 - \ell_k) - u(60 - d), 0\},$$

$$\eta_L(A) = -\frac{1}{K}\sum_{k=1}^{K} \max\{u(\ell_k) - u(a), 0\} + \max\{u(60 - \ell_k) - u(60 - a), 0\}.$$

Each formula includes the sums over two max operators. The first corresponds to the dissatisfactions of the decision maker due to all other allocations in L, and the second corresponds to the similar dissatisfactions of the recipient. The division by K, the number of allocations in L, makes sure that the normative valences can be compared when the size of L is changing. So, if the decision maker takes into account the meta-context created by the experiences of past and currently available allocations described by L, she will perceive the normative valences of D and A as $\eta_L(D)$ and $\eta_L(A)$ instead of $\eta_c(D)$ and $\eta_c(A)$ as was the case in the context model. To estimate this *meta-context model*, we use $\eta_L(D) - \eta_L(A)$ as an independent variable in the regression (2).

To illustrate how the meta-context norm functions are computed, consider Figure 2. Suppose that $t = 2$ and that in this period the dictator is choosing between $D_t = (30, 30)$ and $A_t = (50, 10)$. In the previous period, $t - 1 = 1$, her choice was between allocations $D_{t-1} = (40, 20)$ and $A_{t-1} = (60, 0)$. Thus, the meta-context consists of four allocations: $(60, 0)$, $(50, 10)$, $(40, 20)$, and $(30, 30)$. The dashed line on Figure 2 shows the normative valences computed for all four of them in this meta-context. We can see that the values of $\eta_L(D_t; 0.9)$ and $\eta_L(A_t; 0.9)$ are now in the opposite relationship than in the context model. In this meta-context, choosing A_t is more appropriate than choosing D_t. Now, suppose that 1) $t = 3$; 2) that the choices in the current period and in the one before that are as above; and 3) that the choice in the first period $t - 2$ is between allocations $A_{t-2} = (20, 40)$ and $D_{t-2} = (10, 50)$. In period 3, the meta-context consists

[7] Here, by "experienced" we mean that a subject chose between these allocations in the past, though her specific choice is irrelevant.

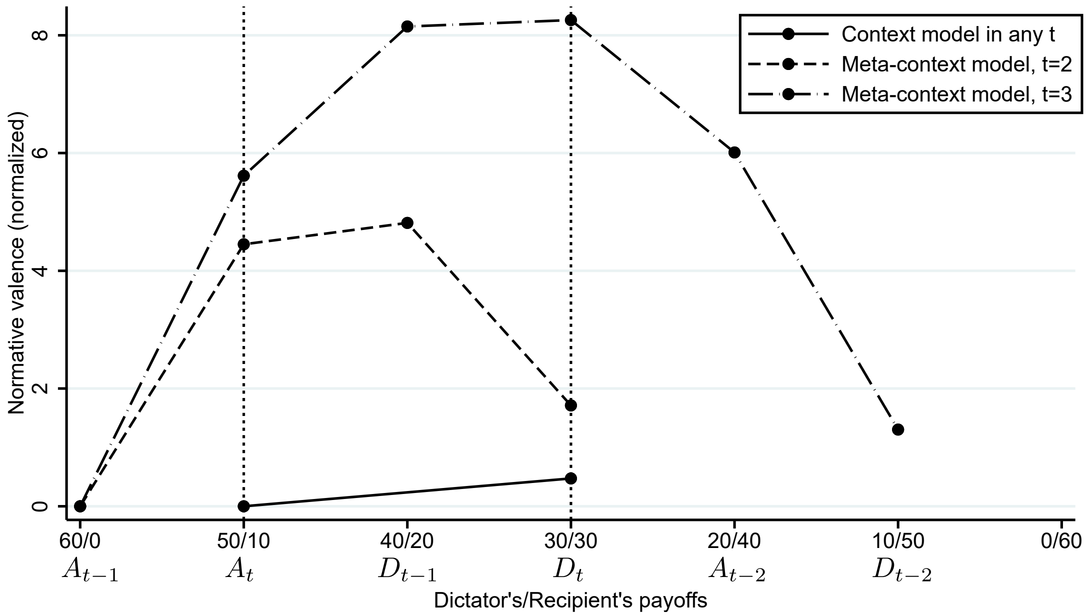

Figure 2: The context and the two meta-contexts of the choice between $(50, 10)$ and $(30, 30)$.

of 6 allocations: $(60, 0)$, $(50, 10)$, $(40, 20)$, $(30, 30)$, $(20, 40)$, and $(10, 50)$. The dot-dashed line shows the normative valences in this meta-context. As we can see, the difference $\eta_L(D_t; 0.9) - \eta_L(A_t; 0.9)$ is positive again, as in the context model, but larger.

To calculate the normative valences of the allocations in period t that include meta-context, we consider subject i's unique experience of past allocations prior to and including period t, which gives us $2t$ allocations in the set L_{it}. For each subject i and each time period t we use the formula above to compute $\eta_{L_{it}}(D_{it}; \alpha) - \eta_{L_{it}}(A_{it}; \alpha)$ and use this as an independent variable in the regression (2). We do this for several values of α, to determine which one fits our data the best (see Section 4).

4 Results

First, we present our main results related to the context-dependence in elicited beliefs and the regressions described above. Then, in the later sections, we discuss the auxiliary and supporting results concerning the relationship between meta-context and the rule-following propensity.

4.1 Main Result

We start with the analysis of normative beliefs elicited in the KW task. The question we ask is whether or not the elicited beliefs reflect the context dependence suggested by the context model in Section 3. Specifically, we ask if the beliefs about the social appropriateness of an allocation depend on the choice contexts in which it is presented (some other paired allocations). For each allocation, that was presented multiple times in the context of various other allocations, we run sign-rank tests to determine if, for some of these presented alternatives, two appropriateness ratings of a given allocation are different. There are 10 allocations that are repeated two, three, or four times in different contexts. Out of 25 sign-rank tests thus obtained, there is only one that is significant at 5% level (three at 10% level), which does not survive any type of correction for multiple comparisons. From this we conclude that the participants do not perceive the appropriateness of an allocation as different depending on the context in which it is presented, as the context model would predict.

There are two possible reasons why participants' beliefs might not be sensitive to the context. One possibility is that they actually perceive the appropriateness of the allocations in a context-independent way, which could be some kind of "outcome-based" normative beliefs. Another possibility is that participants take meta-context into account, as suggested in Section 3, and report the beliefs that come from it. This is plausible since the KW task comes after the mini-DGs, so, at the time of norm elicitation, all participants have already seen all possible allocations.

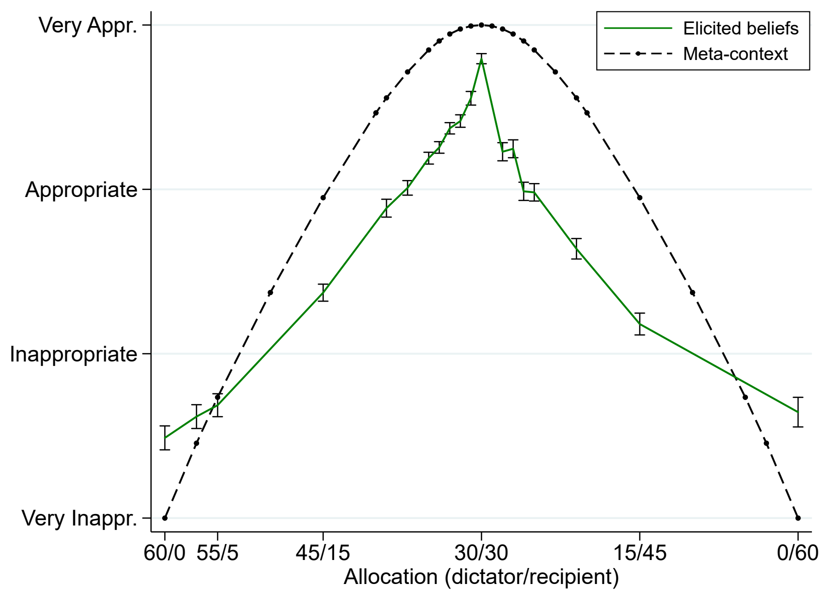

Figure 3: Average appropriateness ratings across allocations and the norm function predicted by the model with all possible allocations taken into account. Error bars are ±1 SE.

To test these ideas, we plot the averages of the elicited beliefs in Figure 3. We average over all occurrences of each allocation in all contexts, since we do not find evidence that the ratings are statistically different. We see that the shape of average beliefs is very similar to that in the standard Dictator game (Krupka and Weber, 2013), which may support the idea that the beliefs are context-independent. Figure 3 also shows the theoretical *norm function* obtained from the meta-context that includes all possible allocations in the experiment (marked by black circles). Given the similarities in shape, the elicited beliefs might be reflecting the meta-context norm function. From this analysis, we cannot tell whether the beliefs are context-independent or whether we are seeing a somewhat distorted version of the meta-context norm function. We also cannot tell whether or not these beliefs are driving the participants' choices. In order to answer these questions, we estimate the three regressions discussed in Section 3.

$\alpha = 0.9$	Elicited norms	Context model	Meta-context model
β	1.368***	3.972***	0.818***
	(0.096)	(0.137)	(0.101)
ϕ	1.262***	48.675***	9.556***
	(0.037)	(1.174)	(0.224)
constant	−3.014***	−2.788***	−2.870***
	(0.249)	(0.246)	(0.247)
N observations	30,212	30,212	30,212
N independent	166	166	166

Table 1: Random-effects logit regressions of the probability of choosing disadvantageous allocation as dependent on material payoffs and a model of the norm function (different for each column). In all models we use $\alpha = 0.9$. Standard errors in parenthesis. *** - $p < 0.001$.

Table 1 shows the estimates of the coefficients in the three regressions. Appendix A describes in detail how the normative valences are interpolated from the elicited norms in the first regression. In the other two regressions we use the independent variables constructed exactly as described in Section 3 with $\alpha = 0.9$ (the reason for using this specific value is explained below). We see that all coefficients in all three models are highly significant.[8] In order to understand which model fits our data the best we compute AIC and BIC. The results are shown in Table 2.

	df	AIC		BIC	
		value	Δ	value	Δ
Meta-context model	4	18,348		18,381	
Context model	4	18,482	134	18,515	134
Elicited norms model	4	19,367	1,019	19,400	1,019

Table 2: Model fit comparisons of the three models. "df" stands for the degrees of freedom. The models are sorted by the value of BIC.

We can see that the meta-context model is by far the best at explaining our data. Both AIC and BIC are the lowest with a large difference of 134 from the next best (context) model.[9] The regression with elicited norms does the worst job. The reasons for this are now clear: the best fit of the

[8] Overall, subjects in our experiment are rather pro-social (defined as choosing disadvantageous allocations). Only around 17% of subjects have chosen selfishly in all 182 problems. The rest chose pro-socially in around half of the problems on average. Figure 6 in Appendix C shows the histogram of proportions of pro-social choices.

[9] When judging whether one model explains the data better than the other in terms of AIC and BIC, a difference in the values of these criteria higher than 6 is usually taken as a sign of superiority of one model over the other. Our smallest difference is 134, way beyond this threshold.

meta-context model suggests that our subjects do indeed perceive the meta-context of each decision and choose in accordance with it, which in its turn supports the idea that the elicited norms shown in Figure 3 also come from the meta-context and that the seeming context-independence of the elicited beliefs is actually the perceived meta-context and not some outcome-based beliefs independent of it.

Finally, in Appendix B we show the extended information criteria table that compares several other model specifications with different α and different ways of including past allocations. Specifically, it is shown that the meta-context model with $\alpha = 0.9$ fits our data better than the meta-context models with higher or lower values of α. The elicited norms and context models are also estimated with different values of α. In addition, we check the discounted meta-context model in which past dissatisfactions are given lower weight the farther in the past they are; the renormalized meta-context model, in which norm functions are normalized to $[-1, 1]$ in each period; and lag-1 meta-context model where only current period t and period $t-1$ are assumed to form the meta-context. All these alternative specifications are inferior to the meta-context model in Table 2 in terms of AIC and BIC, which makes it the best model we could find.

Result 1. *The meta-context model with $\alpha = 0.9$ fits our data the best.*

4.2 Meta-Context and the Propensity to Follow Norms

The result in the previous section has demonstrated that the meta-context model is the best at explaining our data. However, the analysis above does not tell us anything about the individual characteristics that make subjects less or more prone to take meta-context into account. According to the theory of KV, meta-context enters the utility through the injunctive norm function, so norm-following subjects should be more sensitive to it than subjects who disregard the norms. We test this idea by utilizing the individual rule-following propensities that we elicited in the Rule-Following task (see Section 2.1).

As Kimbrough and Vostroknutov (2018) have shown, the proportion of balls that subjects place in the blue bucket—which serves as an estimate of how much they choose to follow an arbitrary costly rule—correlates well with pro-social behavior.[10] Figure 7 in Appendix C shows the distribution of the proportions of balls in the blue bucket in our data. It looks exactly the same as similar distributions reported in Kimbrough and Vostroknutov (2018).

In order to check if sensitivity to meta-context has normative nature, we use this individual characteristic—that can be thought of as a proxy for ϕ_i in the utility specification (1)—in the regression analysis. We define two groups of participants, "rule-followers" and "rule-breakers," depending on whether their proportion of balls in the blue bucket is above or below the median,

[10] This measure has also been successfully used in other experiments (e.g., Thomsson and Vostroknutov, 2017; Gürdal et al., 2020). See also Kimbrough and Vostroknutov (2016).

and interact the resulting categorical variable rule-follower with all independent variables in the regressions presented in Table 1.

$\alpha = 0.9$	Elicited norms	Context model	Meta-context model
rule-follower	1.696***	1.773***	1.762***
	(0.472)	(0.471)	(0.472)
β	1.386***	3.962***	1.000***
	(0.160)	(0.232)	(0.171)
ϕ	0.963***	44.762***	8.562***
	(0.060)	(1.909)	(0.366)
rule-follower$\times\beta$	−0.043	0.029	−0.289
	(0.199)	(0.288)	(0.212)
rule-follower$\times\phi$	0.473***	6.362**	1.574***
	(0.077)	(2.422)	(0.462)
constant	−3.844***	−3.686***	−3.761***
	(0.344)	(0.343)	(0.344)
N observations	30,212	30,212	30,212
N independent	166	166	166

Table 3: Random-effects logit regressions with rule-followers. In all models we use $\alpha = 0.9$. Standard errors in parenthesis. ** - $p < 0.01$; *** - $p < 0.001$.

Table 3 presents the results. As expected, we find that being a rule-follower increases the overall probability of choosing a disadvantageous allocation in all three models (the coefficients on rule-follower). More importantly, we see that the interaction rule-follower$\times\phi$ is positive and significant in all models. This shows that the choices of subjects with higher propensity to follow rules are more sensitive to context and meta-context, thus corroborating the hypothesis that they are related to injunctive norms.

Result 2. *Sensitivity to context and meta-context is related to propensity to follow norms.*

4.3 Sub-Populations

In this section, we address the final important question related to the sensitivity to meta-context. The regressions presented above show that all three models that we consider (elicited norms, context, and meta-context) have significant coefficients on the norm function. The fact that the meta-context model fits our data the best does not guarantee, however, that the three regressions capture the behavior of the same subjects. Hypothetically, it is possible that some subjects in our sample behave in accordance with the elicited norms model, some others in accordance with the context model, and yet others in accordance with the meta-context model.

To understand whether such sub-populations exist, or whether, alternatively, all three models explain the behavior of the same subjects, we estimate individual coefficients β_i and ϕ_i for each subject i. In order to do that, we run three OLS regressions, as those in Table 1, only with all independent variables interacted with the categorical variables representing the choices of each subject.[11] This gives us three sets of 166 coefficients corresponding to three model specifications.

If it is the case that the three models explain the behavior of different sub-populations of subjects, then we should observe negative correlations between the individual coefficients ϕ_i coming from the three models. If, however, all three models fit the behavior of the same subjects, then we should observe positive correlations between individual coefficients ϕ_i coming from different models.

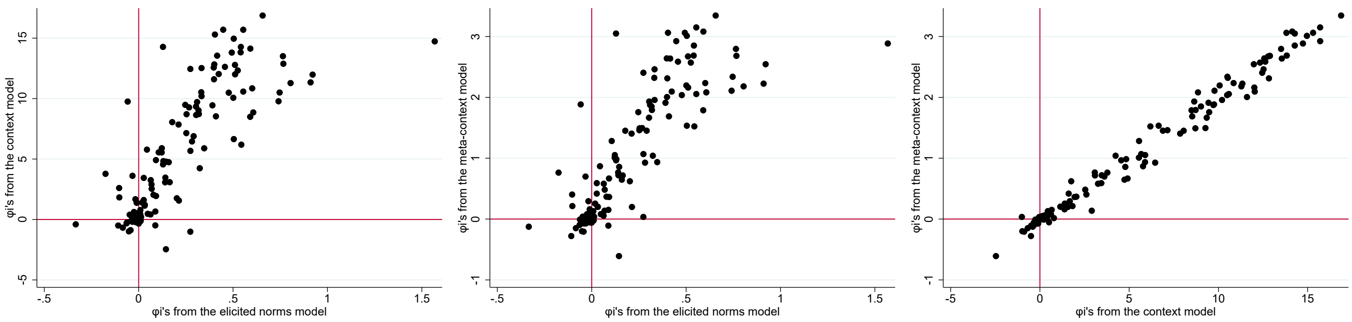

Figure 4: Scatter plots of the individual coefficients ϕ_i from the three models.

Figure 4 shows the scatter plots of individual coefficients ϕ_i coming from the three models. The graphs illustrate that the individual coefficients are highly positively correlated. The rightmost graph, for example, shows a remarkable match between coefficients from the context and the meta-context models (Spearman's $\rho = 0.97$, $p < 0.0001$). Similar, but somewhat noisier connection exists between the individual coefficients from the elicited norms model and the other two (the leftmost and the middle graphs). The Spearman's rank correlations for these two graphs are also high (0.79 and 0.83 respectively, $p < 0.0001$). This unambiguously shows that the three models capture the same behavior in the whole population of subjects. So, we can conclude that there are not specific sub-populations responsive to different models.

Finally, it is worth noting that all three sets of coefficients ϕ_i are correlated with the rule-following propensity elicited in the Rule-Following task. The three Spearman's rank correlations are 0.36 (the elicited norms model, $p < 0.0001$); 0.32 (the context model, $p < 0.0001$); and 0.34 (the meta-context model, $p < 0.0001$). This finding provides strong support to the theory of KV and shows that the behavior captured by the three models is related to adherence to norms.

Result 3. *The three models reflect the same norm-abiding behavior in the whole sample. Correlation of individual coefficients ϕ_i with the propensity to follow norms measured in the RF task corroborates the model of KV.*

[11] It is impractical to use logit for the estimation of individual coefficients. As is often the case, individual logits do not converge for many subjects due to the absence of enough variance in their responses (e.g., some subjects almost always choose advantageous allocations). OLS regression does not suffer from this caveat.

5 Discussion

Inducing meta-context beliefs is easy. The results of our experiment show that most of our subjects are sensitive to the meta-context of their choices, and that this sensitivity is synonymous with the sensitivity to following norms in general. So, subjects who are willing to follow norms are exactly those who are mostly affected by the meta-context. Given that each subject observes an individual randomized sequence of 182 mini-Dictator decisions, we can be sure that this result is not an artifact of the specific sequence of decision problems. Moreover, this result cannot be attributed to any conventional form of (social) learning since subjects never receive any feedback that they are uncertain about: they only make choices between two allocations at a time with known consequences of their choice for both players involved. Overall, we conclude that inducing meta-context norms is "easy," which suggests that the policies directed at changing normative incentives by inducing meta-context beliefs can potentially be useful. The examples of successful interventions designed along these lines are presented in Beaman et al. (2012) and Breda et al. (2020).

Experiencing versus Knowing. One important question that our experiment raises is how much "experience" with unavailable allocations should people have in order to "include" them in the meta-context. We show that the experience with being able to determine which allocation will be implemented does the job, though this is a rather high level of involvement. At the same time, it is clear that simply *knowing* how people live in other countries (e.g., by reading a newspaper) is not enough to induce meta-context beliefs since everyone is roughly aware of the conditions in different locations, but this knowledge does not seem to change people's moral attitudes too much.[12] More research is needed to exactly determine what degree of experience with other potentially possible allocations is needed to induce meta-context beliefs.

Stability of closed societies. Consider a society that consists of two strata: the ruling elite and the general population. Assume that the elite determines how a "pie" (e.g., the money collected from taxes) is divided and assume that there is a norm that the elite gets 90% of the money and the rest of the population 10%. As long as the population does not question the norm, the resulting 90/10 division will be stable, because the population considers the division socially appropriate (see the left panel of Figure 1) and the elite does not want to change anything given that it enjoys very high profits already. Thus, in the absence of the flow of information about the meta-context, our model suggests that such arrangements can be stable. Our result that inducing meta-context beliefs is easy also suggests that the elites in such societies should create information barriers in order to keep their population uninformed and to maintain the existing status quo. With the spread of internet, we see more and more examples of such restrictions in the developing world (e.g., the great firewall of China). Our framework points out that the

[12] Though, there is evidence that internet has positive effect on democratization (Boulianne, 2020).

emergence of such information barriers signifies an attempt by the elites to keep the population's choice context restricted and that the elites might fear information spread due to meta-context effects that it entails.

Other signs of restricted context. Many countries with multi-national and heterogeneous populations, including the US and countries in the European Union, suffer from economic and political exclusion of certain social groups (e.g., African-Americans or immigrants from North Africa). The slow process of assimilation of these groups reflects the possibility that they accept their subordinate position in the society (base their moral judgements on a restricted context). Thus, inducing meta-context beliefs (by encouraging them to pursue careers they deem infeasible) might be a way to help these groups to integrate into economic and political process.

Unfalsifiability. It may seem that our model of meta-context effects is unfalsifiable since any behavior or beliefs can be "explained" by the presence of some specific meta-context. We do not deny that this is a possibility. However, our experiment clearly demonstrates that there are *some* meta-context effects, even if our model is not capturing them correctly. This is evident from the presence of a relationship between the allocations experienced in the past and the current choice showcased in our regression analysis. We believe that in the world where meta-context effects exist, it is better to have some model and continue research that determines what can and what cannot be part of the meta-context, than to have no model and stop all attempts at explaining meta-context effects on the grounds that any such theory can potentially explain anything.

Connection to regret avoidance. It is interesting to note that the model of KV, when restricted to one player, becomes a model of "regret avoidance," where the notion of regret is extended from the payoffs that can be received in the current decision problem (a standard definition) to payoffs that have been experienced in the past. The experiment of Fioretti et al. (2021) provides evidence of such regret avoidance. In the experiment, subjects observe the prices of an asset as time unfolds and choose when to sell it. The results show the effect of past regret on the selling decisions: subjects hold the asset longer than it is optimal due to the effect of high prices observed in the past. This can be seen as the effect of meta-context in individual decision problems. It is, therefore, possible that regret of the kind detected in Fioretti et al. (2021) is the building block on which moral judgements are based, as suggested by KV.[13] More research is needed to identify the connection between social behavior and regret avoidance.

6 Conclusion

In this study we show experimentally that moral choices depend on the allocations of payoffs experienced in the past that are not necessarily a part of the decision problem at hand (the meta-

[13] To understand what is socially appropriate and what is not, people compute how much "regret" other players might have in different outcomes.

context). We use the theory of injunctive norms by Kimbrough and Vostroknutov (2020) to show why and how these allocations enter "moral calculus." The idea that moral judgements depend on the beliefs about possible counterfactual allocations suggests that discriminated groups, who base their decisions on a restricted context (e.g., girls in "traditional" cultures think that pursuing a scientific career is not for them), will have weak incentives to change their situation since from the perspective of the restricted context nothing wrong is taking place. Our findings show that it is easy to induce meta-context beliefs. Thus, a new way to fight discrimination might be to create additional normative incentives through meta-context.

References

Akos, P., Lambie, G. W., Milsom, A., and Gilbert, K. (2007). Early adolescents' aspirations and academic tracking: An exploratory investigation. *Professional School Counseling*, 11(1):57–64.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *Science*, 335(6068):582–586.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Boulianne, S. (2020). Twenty years of digital media effects on civic and political participation. *Communication research*, 47(7):947–966.

Breda, T., Grenet, J., Monnet, M., and van Effenterre, C. (2020). Do female role models reduce the gender gap in science? Evidence from french high schools. Technical report, Institute of Labor Economics (IZA).

Ceci, S. J. and Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8):3157–3162.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, 106(6):1691–1730.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American economic review*, 94(4):857–869.

Faris, D. (2013). *Dissent and revolution in a digital age: Social media, blogging and activism in Egypt*. Bloomsbury Publishing.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458–468.

Fioretti, M., Vostroknutov, A., and Coricelli, G. (2021). Dynamic regret avoidance. *American Economic Journal: Microeconomics*, forthcoming.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Greely, B. (2020). IMF chief Kristalina Georgieva: The fund needs to have a big bazooka. *Financial Times*.

Gürdal, M. Y., Torul, O., and Vostroknutov, A. (2020). Norm compliance, enforcement, and the survival of redistributive institutions. *Journal of Economic Behavior and Organization*, 178:313–326.

Henderson, S. (2003). *Building democracy in contemporary Russia: Western support for grassroots organizations*. Cornell University Press.

Isik, U., Tahir, O. E., Meeter, M., Heymans, M. W., Jansma, E. P., Croiset, G., and Kusurkar, R. A. (2018). Factors influencing academic motivation of ethnic minority students: A review. *Sage Open*, 8(2):1–23.

Kessler, J. B. and Leider, S. (2012). Norms and contracting. *Management Science*, 58(1):62–77.

Kimbrough, E. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of European Economic Association*, 14(3):608–638.

Kimbrough, E. and Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, 168:147–150.

Kimbrough, E. and Vostroknutov, A. (2020). A theory of injunctive norms. mimeo, Chapman University and Maastricht University.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of European Economic Association*, 11(3):495–524.

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90:375–402.

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

McFadden, D. L. (1976). Quantal choice analaysis: A survey. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 363–390. NBER.

Read, D., Loewenstein, G., and Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1–3):171–197.

Rodgers, W. M. (2009). *Handbook on the Economics of Discrimination*. Edward Elgar Publishing.

Steuter, E. (1992). Women against feminism: an examination of feminist social movements and anti-feminist countermovements. *Canadian Review of Sociology/Revue canadienne de sociologie*, 29(3):288–306.

Thomsson, K. and Vostroknutov, A. (2017). Small-world conservatives and rigid liberals: Attitudes towards sharing in self-proclaimed left and right. *Journal of Economic Behavior and Organization*, 135:181–192.

Voicu, M., Voicu, B., and Strapcova, K. (2009). Housework and gender inequality in European countries. *European sociological review*, 25(3):365–377.

Appendix (for online publication)

A Interpolation of the Elicited Beliefs

In this appendix we show how we interpolate the normative valences from the elicited beliefs to later use them in the regression presented in Table 1. Given that the elicited ratings do not depend on the context (see Section 4.1), we use all allocation ratings taken separately to interpolate the appropriateness of all other allocations in the mini-DGs. We estimate the norm function from elicited beliefs as a piece-wise power curve for allocations that give the dictator less or more than the half of the pie, fitted separately for each participant (see Figure 5).

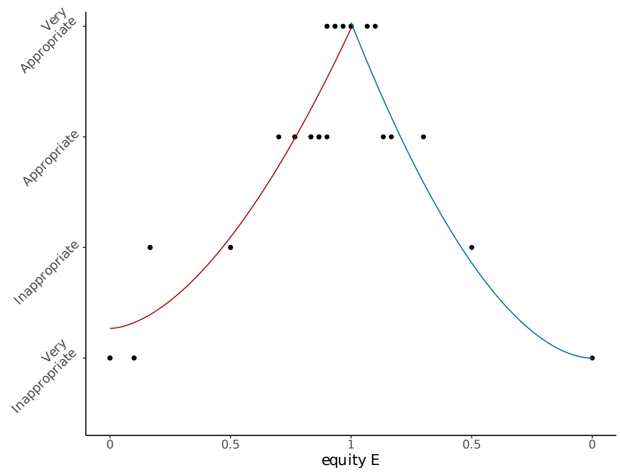

Figure 5: An example of a power function fitting (participant 17).

The red and blue lines of the figure show the model fit and serve as inputs into the regression in Table 1 that uses elicited beliefs. The fitting was done using the following non-linear regression:

$$N(\pi) = \begin{cases} a_1 + b_1 E^{c_1} + \varepsilon_1 & \text{if } \pi \geq 30 \\ a_2 + b_2 E^{c_2} + \varepsilon_2 & \text{if } \pi \leq 30 \end{cases}$$

The parameters a_k, b_k, c_k for $k = 1, 2$ represent the two power curves. The dependent variable $N(\pi)$ stands for the social appropriateness ratings of the 19 allocations used in the KW task. The values of $N(\pi)$ are in the set $\{-1, -\frac{1}{3}, \frac{1}{3}, 1\}$, where -1 stands for "very inappropriate" and 1 for "very appropriate." The independent variable E is a linearly normalized measure that orders allocations by their equality, ranging from 0 (highest inequality, e.g., 60/0 or 0/60) to 1 (complete equality, e.g., 30/30). Errors ε_k are assumed to be normally distributed. A parameter summary is shown in Table 4.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
a_1	–1.18	–1.04	–0.97	–0.69	–0.70	1.09
b_1	–1.29	1.44	1.78	1.52	1.99	2.40
c_1	0.00	0.98	1.49	8.53	2.14	538.38
a_2	–1.14	–1.01	–0.95	–0.58	–0.34	1.08
b_2	–0.86	1.11	1.65	1.44	1.94	2.24
c_2	0.00	1.35	2.29	44.16	6.83	3637.47

Table 4: Summary of the individual interpolation parameters.

B Additional Models

In this appendix we analyze several other model specifications that are different in the way the norm function is computed. Table 5 shows the AIC and BIC.

	α	λ	df	AIC	BIC
Meta-context model	0.90		4	18,348	18,381
Meta-context model	0.80		4	18,348	18,381
Meta-context model	0.95		4	18,353	18,386
Meta-context model	0.98		4	18,358	18,391
Meta-context model	1.00		4	18,361	18,395
Context model	0.98		4	18,460	18,493
Context model	0.95		4	18,467	18,501
Context model	0.90		4	18,482	18,515
Context model	0.80		4	18,518	18,551
Discounted meta-context model	0.90	0.90	4	18,548	18,581
Discounted meta-context model	0.90	0.80	4	18,641	18,675
Renormalized meta-context model	0.90		4	18,647	18,680
Elicited norms model	0.80		4	19,348	19,381
Elicited norms model	0.90		4	19,367	19,400
Elicited norms model	0.95		4	19,367	19,409
Elicited norms model	0.98		4	19,381	19,415
Elicited norms model	1.00		4	19,385	19,418
Lag-1 meta-context model	0.80		4	19,572	19,606
Lag-1 meta-context model	0.90		4	19,713	19,746

Table 5: Model fit comparisons. "df" stands for the degrees of freedom. The models are sorted by the value of BIC.

Most models in the table are the same as in Table 2 except for different parameter α that enters the consumption utility ($u(x) = x^{\alpha}$) and consequently the norm functions in the context and the meta-context models. In Table 5, we observe that for other values of α the model fit is worse than for the meta-context model with $\alpha = 0.9$.[1] In the elicited norms model, α enters only through the consumption utility.

In addition to changing α, we have also checked three different specifications of the meta-context model. In the discounted meta-context model the computations of the (meta-context) normative valences are the same as in Section 3 except that instead of taking the average over all dissatisfactions, we discount them differently depending on how far away in the past they are. So, the dissatisfactions from the current allocations have weight 1, the dissatisfactions coming from the previous period allocations are discounted with the parameter λ, allocations from two periods ago with λ^2, etc. (see Table 5). As we can see the fit of these models is much worse than the fit of the original meta-context model where all past dissatisfactions have the same weight. This is an important finding that tells us that subjects literally consider all other *possible* allocations, regardless of when they took place in the past. This supports the framework of KV in which this assumption is "built-in."

The renormalized meta-context model uses the same calculations as the usual meta-context model except that in each period the normative valences are renormalized to the interval $[-1, 1]$, as suggested in KV. This model does not do a very good job at explaining the data. This may seem as a falsification of this specific assumption in the KV model. However, we would like to point out that KV's setup is designed

[1] The AIC and BIC for the meta-context model with $\alpha = 0.8$ are slightly higher than for the the case with $\alpha = 0.9$.

to work with a *single* context. The renormalization to $[-1, 1]$ in this case does not change anything mathematically, but makes derivations easier. KV, however, do not talk about whether this renormalization should be used for situations in which the context and the meta-context are changing.

Finally, we consider the lag-1 meta-context model, which is the same as the usual meta-context model except that only four allocations are taken into account: the current allocations plus the two allocations from the previous period. This model fares the worst, even worse than the elicited norms model.

All these results considered together suggest that all past allocations matter for the computation of meta-context. Moreover, they all equally important regardless of when they have occurred in the past.

C Additional Graphs

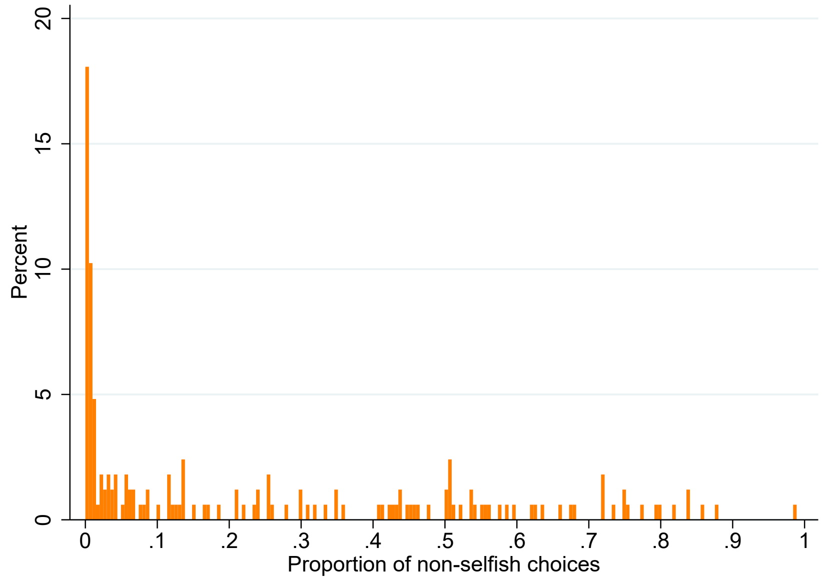

Figure 6: Histogram of the proportion of non-selfish choices for each subject (166 observations).

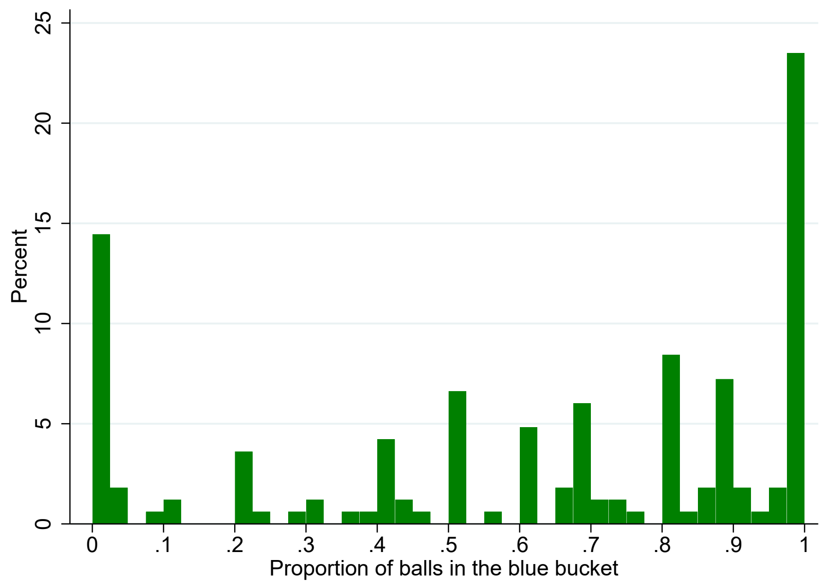

Figure 7: Histogram of the number of balls in the blue bucket.

D Details of the Design

Allocation							
A	B	A	B	A	B	A	B
3/57	60/0	30/30	33/27	37/23	26/34	50/10	31/29
5/55	55/5	30/30	35/25	37/23	27/33	50/10	34/26
5/55	60/0	30/30	39/21	37/23	30/30	50/10	35/25
10/50	50/10	30/30	45/15	37/23	31/29	50/10	40/20
10/50	57/3	30/30	55/5	37/23	34/26	50/10	45/15
15/45	50/10	30/30	60/0	37/23	35/25	50/10	57/3
15/45	57/3	31/29	29/31	37/23	40/20	55/5	10/50
20/40	45/15	31/29	33/27	37/23	50/10	55/5	15/45
20/40	55/5	31/29	35/25	37/23	57/3	55/5	23/37
20/40	60/0	31/29	39/21	39/21	23/37	55/5	25/35
21/39	39/21	31/29	45/15	39/21	25/35	55/5	28/32
21/39	45/15	31/29	55/5	39/21	28/32	55/5	29/31
21/39	55/5	31/29	60/0	39/21	29/31	55/5	32/28
21/39	60/0	32/28	30/30	39/21	32/28	55/5	33/27
23/37	37/23	32/28	31/29	39/21	33/27	55/5	37/23
23/37	40/20	32/28	34/26	39/21	37/23	55/5	39/21
23/37	50/10	32/28	37/23	39/21	40/20	55/5	50/10
23/37	57/3	32/28	40/20	39/21	50/10	55/5	57/3
25/35	37/23	32/28	50/10	39/21	57/3	57/3	3/57
25/35	40/20	32/28	57/3	40/20	20/40	57/3	5/55
25/35	50/10	33/27	28/32	40/20	21/39	57/3	20/40
25/35	57/3	33/27	29/31	40/20	26/34	57/3	21/39
26/34	35/25	33/27	32/28	40/20	27/33	57/3	26/34
26/34	39/21	33/27	34/26	40/20	30/30	57/3	27/33
26/34	45/15	33/27	37/23	40/20	31/29	57/3	30/30
26/34	55/5	33/27	40/20	40/20	34/26	57/3	31/29
26/34	60/0	33/27	50/10	40/20	35/25	57/3	34/26
27/33	33/27	33/27	57/3	40/20	45/15	57/3	35/25
27/33	35/25	34/26	26/34	40/20	55/5	57/3	40/20
27/33	39/21	34/26	27/33	40/20	60/0	57/3	45/15
27/33	45/15	34/26	30/30	45/15	15/45	57/3	60/0
27/33	55/5	34/26	31/29	45/15	23/37	60/0	0/60
27/33	60/0	34/26	35/25	45/15	25/35	60/0	10/50
28/32	32/28	34/26	39/21	45/15	28/32	60/0	15/45
28/32	34/26	34/26	45/15	45/15	29/31	60/0	23/37
28/32	37/23	34/26	55/5	45/15	32/28	60/0	25/35
28/32	40/20	34/26	60/0	45/15	33/27	60/0	28/32
28/32	50/10	35/25	25/35	45/15	37/23	60/0	29/31
28/32	57/3	35/25	28/32	45/15	39/21	60/0	32/28
29/31	32/28	35/25	29/31	45/15	55/5	60/0	33/27
29/31	34/26	35/25	32/28	45/15	60/0	60/0	37/23
29/31	37/23	35/25	33/27	50/10	20/40	60/0	39/21
29/31	40/20	35/25	39/21	50/10	21/39	60/0	50/10
29/31	50/10	35/25	45/15	50/10	26/34	60/0	55/5
29/31	57/3	35/25	55/5	50/10	27/33		
30/30	31/29	35/25	60/0	50/10	30/30		

Table 6: Mini-DGs used in the experiment (dictator's/recipients' payoffs).

Allocation		A			B		
A	B	mean	SE	median	mean	SE	median
0/60	60/0	-0.570	0.060	-1.000	-0.622	0.057	-1.000
15/45	55/5	-0.213	0.044	-0.333	-0.546	0.050	-1.000
21/39	57/3	0.092	0.042	0.333	-0.610	0.051	-1.000
25/35	55/5	0.321	0.035	0.333	-0.554	0.049	-1.000
26/34	35/25	0.325	0.037	0.333	0.482	0.026	0.333
27/33	33/27	0.498	0.036	0.333	0.606	0.027	0.333
28/32	37/23	0.486	0.037	0.333	0.337	0.032	0.333
30/30	34/26	0.847	0.028	1.000	0.502	0.028	0.333
30/30	39/21	0.867	0.025	1.000	0.257	0.036	0.333
30/30	60/0	0.859	0.027	1.000	-0.719	0.050	-1.000
30/30	45/15	0.880	0.024	1.000	-0.096	0.039	-0.333
31/29	37/23	0.703	0.028	1.000	0.341	0.033	0.333
32/28	35/25	0.606	0.030	0.333	0.438	0.031	0.333
32/28	57/3	0.614	0.032	0.333	-0.566	0.055	-1.000
33/27	34/26	0.550	0.027	0.333	0.482	0.030	0.333
33/27	45/15	0.586	0.031	0.333	-0.092	0.039	-0.333
55/5	34/26	-0.526	0.051	-1.000	0.526	0.030	0.333
60/0	45/15	-0.683	0.053	-1.000	-0.068	0.039	-0.333

Table 7: Items used in the norm elicitation task and the rating summary.

E Instructions

GENERAL INFORMATION
Dear participant,
You will now take part in an experiment consisting of three decision-making tasks and a questionnaire. The amount of money that you will earn during the experiment will be paid to you in cash at the end of the experiment. Your earnings for each part of the experiment will be shown to you only at the end of the experiment. We ask you to please not talk to other participants during the experiment. If you have any questions, please raise your hand: one of the experimenters will come to your station and will answer your questions.

PART 1: FIRST DECISION-MAKING TASK

During the first part of the experiment, you will decide how to distribute 50 balls between two buckets. Your task is to drag-and-drop each ball, one at a time, into one of the two buckets in front of you. The balls will appear at the center of your screen, and you can allocate each ball by clicking on it and dragging it in the bucket of your choice. For each ball placed in the blue bucket, you will earn €0.05, and for each ball placed in the yellow bucket, you will earn €0.10. The rule is to place the balls in the blue bucket. Once you start the experiment, you will have 5 minutes to place the balls in the buckets. When you are finished, please wait silently until the end of the 5 minutes. The payment will be based on your choices. This is the end of the instructions of part 1. If you have questions, please raise your hand: an experimenter will come to your station and will answer your questions in private. Otherwise, please wait silently that the experiment starts. Press OK when you are ready

PART 2: SECOND DECISION-MAKING TASK

In this part of the experiment you will be shown several possible splits of 60 points between yourself and another participant. The amount of points on the black background is the part of 60 points that would be given to you, while the amount of points on the white background is the part of 60 points that would be given to another randomly chosen participant in this experiment. For each split of points that you will be shown, we ask you to choose which one of the two do you prefer. To choose your preferred option, click on it: the option will be highlighted and a CONFIRM button will appear on top. Press the CONFIRM button to confirm your choice and to move to the next split. Remember: you can change your choice as many times as you want as long as you do not press the CONFIRM button.

PART 2: PARTICIPANT SELECTION AND PAYMENT

All other participants in the experiment will complete the same task with the same splits at the same time as you. At the end of the experiment, half of participants will be selected randomly. Each of the selected participants will be paired and will split points with one of the non-selected participants. Therefore, if you are selected, one of your choices will be implemented and the points will be split with the other participant. If instead you are not selected, you will receive points from one of the choices of a selected participant. If you are not selected your choices will not be used, only the other participant's choices will matter. Both the specific choice and the pairing will be random. Each point corresponds to €0.10.

PART 2: ANONIMITY

Please note: all your choices are anonymous, neither you nor other participants will ever be able to identify the actions of the other participants, or to identity the person you will be paired with. This is the end of the instructions of part 2. If you have questions, please raise your hand: an experimenter will come to your station and will answer your question in private. Otherwise, press ENTER THE EXPERIMENT when you are ready.

PART 3: THIRD DECISION-MAKING TASK

Consider the following situation: a person is asked to complete the same task that you have just completed, with the same rules. You will be shown some of the splits that you have already seen, where the person had to choose his/her preferred combination. For each pair of combinations, you have to rate how socially appropriate or inappropriate it is that this person chooses the left combination, and how socially appropriate or inappropriate it is that this person chooses the right combination. By socially appropriate we mean a choice that most people would find "correct" or "right" thing to do. In other words, if this person were to choose a socially inappropriate combination, then someone could become upset with this person. Along with the splits, on the screen you will be presented with a table in which you rate the degree of appropriateness of each combination. To select the level of social appropriateness of a combination, click on the option that you consider correct (among: very inappropriate, rather inappropriate, rather appropriate, and very appropriate).

PART 3: PAYMENT

At the end of the experiment, one of the 'left' or 'right' choices will be randomly selected among the pairs of combinations that you were asked to rate. For this choice, it will be determined what answer (very inappropriate, rather inappropriate, rather appropriate, very appropriate) was most frequently selected by all other participants. If your choice concides with that of the majority you will be paid €5. Remember: do not rate following what you consider 'correct' or 'right' yourself, but rather how you think most people would rate the choice. This is the end of the instructions of part 3. If you have questions, please ask by raising your hand: an experimenter will come to your station and will answer you in private. Press OK when you are ready.