Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 39-2020

Elina Khachatryan and Christoph Bühren

Letting off Steam! Experimental Evidence on Inappropriate Punishment

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Letting off Steam! Experimental Evidence on Inappropriate Punishment

ELINA KHACHATRYAN
University of Kassel
Address: Nora-Platiel-Straße 4, 34127 Kassel, Germany
Telephone: +49 561 804-2327
Email: elina.khachatryan@uni-kassel.de

CHRISTOPH BÜHREN
Clausthal University of Technology
Address: Julius-Albert Str. 2, 38678 Clausthal-Zellerfeld, Germany
Telephone: +49 5323 72-7630
Email: christoph.buehren@tu-clausthal.de

Abstract
Aggression is displaced when provocations cannot be directly retaliated against and when it is redirected towards a target innocent of any wrongdoing. While this phenomenon is widespread, it has not been widely explored in experimental economics. We fill this gap and find that a sizeable proportion of subjects (37%), when treated unfairly, punish co-players who are not at all responsible for the unfairness. When in a disadvantaged position, inequity-aversion seems to be the driving force of punishment, yet when treated fairly, some subjects (17%) exhibit status-seeking behavior. Moreover, students affiliated with an armed forces university are much more likely than regular students to engage in displaced aggression.

JEL: C91; D03; D63.
Keywords: Displaced Aggression; Punishment; Soldiers; Dictator Game, Experiment

Highlights

- Aggression is displaced when it is redirected towards an innocent target.
- When treated unfairly, some punish peers who are not responsible for the unfairness.
- Soldier-students are more likely to use displaced aggression than regular students.

1. Motivation

Economic experimental literature has shown that punishment is sometimes exercised for the emotional satisfaction of the punisher rather than as a disciplining effort directed towards an potential long-term cooperation partner. Fehr and Gächter (2000) and various replications and extensions of their study demonstrate that subjects use costly punishment against co-players that upset them even if a future interaction is not foreseen. Dickinson and Masclet (2015) find that excessive punishment is driven by venting emotions, and Drouvelis and Grosskopf (2016) show that subjects induced (via video clips) with the emotion of anger are harsher punishers.

Our study takes this issue to the next level by looking at the consequences of an unfulfilled desire to punish an anger-provoking co-player. Dollard et al. coined the term *Displaced Aggression* in 1939 already for situations where an innocent bystander is punished when the original provocateur of aggression is out of reach.[1] A real-life example of "being at the wrong place at the wrong time" is investigated by Card and Dahl (2011). Using police reports data of domestic violence incidents during the professional football season in the United States, they show that "unexpected" losses (when an objectively better-rated team loses) lead to a 10% increase in the rate of at-home violence by men against their wives and girlfriends.

To the best of our knowledge, we are first to implement economic experimental methods to study displaced aggression. Our experiment is an adaptation of the Bartling and Fischbacher (2012) four-person game.[2] One of these four players, the dictator, has the incentive to act selfishly but cannot be penalized for her[3] potentially selfish behavior. The dictator can benefit herself and another player at the cost of two remaining players. Given such an advantaged position in the gamethe majority of subjects in the role of the dictator, not surprisingly, abuse their power.

[1] Strangely, popular culture sometimes seems to encourage the displacement of aggression towards inanimate objects. For example, Bushman et al. (1999) report that some self-help books encourage one to punch pillows or break glass to "let off steam." The expression "letting off steam" is not exclusive to the Anglo-Saxon culture. A five-lingual co-author of this paper can think of a colloquialism that expresses the same sentiment in all the five languages she speaks. The evidence of Bushman et al. (1999) points into the opposite direction: Outward aggression towards inanimate objects does not help one to "let off steam," but only "fires one more up."

[2] However, we study a completely different question. Bartling and Fischbacher's (2012) study whether the delegation of the distributive decision to another person also shifts the blame for unfairness to this other person. They find that the decision to delegate is often perceived by the "delegee" as a signal to choose the unfair distribution. And then indeed the "delegees" are more likely to get punished, although the original decision-maker could have directly chosen the fair distribution themselves.

[3] For simplicity, the dictator (Player A) will always be referred to as "she" and all subjects who can be taken advantage of (Player Cs) will always be referred to as "he".

The focus of the paper, however, lies in the aftermath of such an unequal power structure. The potentially disadvantaged subjects are given the possibility to "let off steam," i.e. exercise costly punishment against their non-dictator co-players.

We use this simple two-stage game to address four research questions:

1. Do innocent bystanders pay for the “sins” of unfair impune[4] dictators?
2. Do *unkinder* intentions by the dictator cause more displaced aggression towards innocent bystanders?
3. Do bad intentions cause more displaced aggression than *bad luck*?
4. Does association with the military sub-culture affect the level of displaced aggression towards innocent bystanders?

The association with the military in our experiment is defined by subjects’ real-life self-selection into the military profession in Germany. Wiens et al. (2015) argue that due to military training, soldiers should be more cooperative and should exert higher social responsibility. They find that in dictator, ultimatum, and trust games, soldiers are more altruistic, more cooperative, more trusting, and more trustworthy than regular students (see Michailova and Bühren (2015) for similar results). Furthermore, the experiment on counterterrorism of Mintz et al. (2006) shows that military officers behave differently compared to students: While less than two-thirds of the students recommended interventions against terrorism, more than 90 percent of military officers recommended them. Moreover, military officers as compared to students behave rather as satisficers than maximizers (Simon, 1955).

The fourth question has been inspired by the knowledge of intercultural differences in aggression norms. Through ethnographic studies, Fessler (2006) demonstrates just how wide the range of cultural differences in aggression norms is. The well-being of an individual of the *Utku* society in the arctic region depends heavily on his/her community’s common fishing and hunting efforts. Since interpersonal conflict can be detrimental for the whole community, it becomes understandable why in this culture aggressive behavior is considered “childish” and is ridiculed. It is then fascinating that aggression can be considered respectably “manly” and almost admirable in another culture. The whole cinematic industry of “Westerns” attests to the “manliness” of violent “cultures of honor” and chivalry, characteristic to American South and West. See Sánchez-

[4] Players are impune if they are immune to punishment.

Jankowski (1991) for a similar analysis on culture and aggression in regards to youth gangs in segregated American cities and Mocan (2013) for an analysis of the determinants of vengeful feelings across different countries.

Our study finds that although a large proportion of subjects acts according to standard predictions and does not engage in any costly punishment, a considerable number of subjects exercise displaced aggression and punish their innocent co-players. Subjects associated with the military are more likely to do so. Punishers seem insensitive to the level of *unkindness* by the dictator and do not react differently to bad intentions vs. bad luck. Additionally, we find that about one-fifth of all our subjects exhibits status-seeking behavior, meaning punish others to become highest earner in the group, when treated fairly.

Our paper consists of five sections. Section 2 describes our design and particularities of our subject pools. Section 3 introduces two economic behavioral models that compete to predict the behavior of subjects in our experiment. Section 4 describes empirical results and compares the empirics to the theoretical predictions made in Section 3. The last sub-section of Section 4 reports on several additional control treatments we have conducted as a robustness check. Lastly, Section 5 provides a summary of our work and discusses its implications.

2. Data Collection

Overall, 308 subjects participated in our experiment. 200 of the subjects participated in the main two treatments discussed in the following sub-section 2.1. The remaining 108 subjects participated in additional control treatments. Design and empirical findings of these treatments are summarized in Appendix E.[5]

256 subjects were undergraduate students at two higher education institutions in Germany. These participants were recruited in their economics lectures[6] and took part in the experiment at the corresponding institutions' computer lab. Care was taken that each student took part in the experiment only once and in one treatment only. The experiment was computerized using the

[5] We recommend looking at Appendix E after reading the paper.
[6] Kaiser et al. (2018) show that economics students and students from other disciplines punish equally often in a social dilemma game.

software z-Tree (Fischbacher, 2007). The remaining 52 subjects were recruited and participated in the experiment in a field setting. They played a pen-and-paper version of the game.

To ensure that all subjects carefully read and understood the instructions, they were required to answer several comprehension questions. In each session, the experimenters controlled all the answers before starting the game. Only very few subjects had any difficulties to correctly answer the comprehension questions. In those rare cases the experimenters showed the answer keys and explained the underlying calculations.

2.1. Experimental Design

To study how people react to unfair treatment in situations where the wrongdoer is impune, we implemented a modified dictator game, inspired by the design of Bartling and Fischbacher (2012). The game has four players: Each group consists of Player A, Player B, and two Player Cs. Player A and Player Cs are active decision-makers, Player B is a passive player.

The game consists of two stages: In Stage 1, Player A, the dictator, makes a binary distributive decision. In Stage 2, using the strategy-eliciting method, Player Cs state their punishment decisions for each of the possible distribution scenarios. Thereafter, one of the Player Cs is randomly selected. This Player's punishment decision is implemented and becomes payoff-relevant to the whole group. Before they make their punishment decisions, however, Player Cs do not know whether they have been selected or their co-playing Player C has been selected. They are, however, well-aware that their punishment decision might be payoff-relevant. Punishment is costly for a Player C (giving up one point to punish by 5 points).

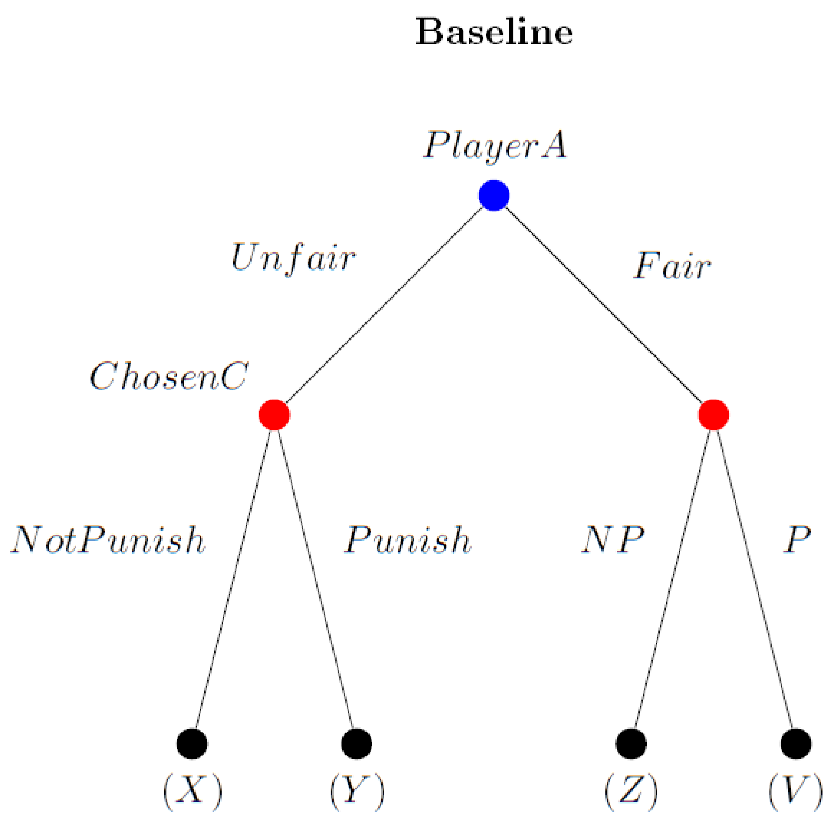
Player B does not make any payoff-relevant decisions in this game. These passive players, however, are asked to state which distribution they would have chosen had they been randomly assigned to the role of Player A (the dictator) – without yet knowing A´s decision.

A very important detail of our design is that Player A, the dictator, cannot be punished. Our design attempts to mimic frustrating conflict situations with strangers who cannot be retaliated against, as for example, frustrating interactions with figures of authority, e.g. a rude police officer. Our research project sets out to explore the aftermath of these situations.

Our experiment has two main treatments. The difference between these treatments occurs in Stage 1: In the *Baseline* treatment, Player A makes a binary choice between an unfair and fair distribution. Our *Chance* treatment is also a binary dictator game. But in this treatment, player A must choose between an unfair distribution and delegation to chance (see also Gurdal et al. 2013). The latter results with equal probabilities in either a fair distribution or an unfair distribution. In both treatments, Player A can directly choose an unfair distribution; only the alternative to this choice is different between the treatments.

Player A begins with 100 points. If Player A directly chooses the unfair distribution, she and Player B in her group receive 45 points each and both Player Cs in the group receive only 5 points each. In *Baseline*, if Player A chooses the fair distribution, she and all the other 3 players in the group receive 25 points each. In *Chance*, payoffs are the same for an unfair distribution regardless of its source: dictator's direct choice or chance. Similarly, a randomly generated fair distribution in *Chance* provides a payoff of 25 points for each of the four players. It is important to note that in *Chance*, Player A cannot directly implement a fair distribution. The *kindest* choice Player A can make here is to delegate the distributive decision to chance; so to say: "give the fair distribution a chance".

In stage 2, Player Cs make the punishment decision before the dictator's distribution is revealed. Player Cs can invest a maximum of 5 points into punishment of Player B and/or the other player C in any of the scenarios, which equals his endowment in the unfair distribution scenarios. Moreover, Player Cs cannot subtract more points from co-players than those already have, meaning that the non-chosen Player C can be punished at most by one point and lose his 5 points in the unfair scenarios.



Payoffs:	***X***	***Y***	***Z***	***V***
A	45	45	25	25
B	45	45-5b	25	25-5b
Other C	5	5-5c	25	25-5c
Chosen C	5	5-b-c	25	25-b-c

Note: "b" = the number of point(s) invested into punishment of Player B and "c" = the number of point(s) invested into punishment of Player B.

Figure 1: *Game Tree* Baseline *treatment*

There are two scenarios in *Baseline* for which Player Cs must provide answers: 1) Player A choosing the unfair distribution and 2) Player A choosing the fair distribution.

Since three different scenarios are possible in *Chance*, Player Cs must state in this case their punishment decisions for each of the three possible scenarios: 1) an unfair distribution directly chosen by Player A, 2) an unfair distribution randomly generated by chance (computer or coin flip, depending on the sessions), and 3) a fair distribution randomly generated by chance (computer or coin flip, depending on the sessions).

Please see the corresponding Figures (1 and 2) for the Game Trees of the *Baseline* and *Chance* treatments, respectively. For simplicity, Player A is always represented by a blue dot, the "chosen" Player C by a red dot, and a random chance draw by a green dot. The payoff table below the *Baseline* Game Tree is relevant also to the payoff structure in the *Chance* treatment.

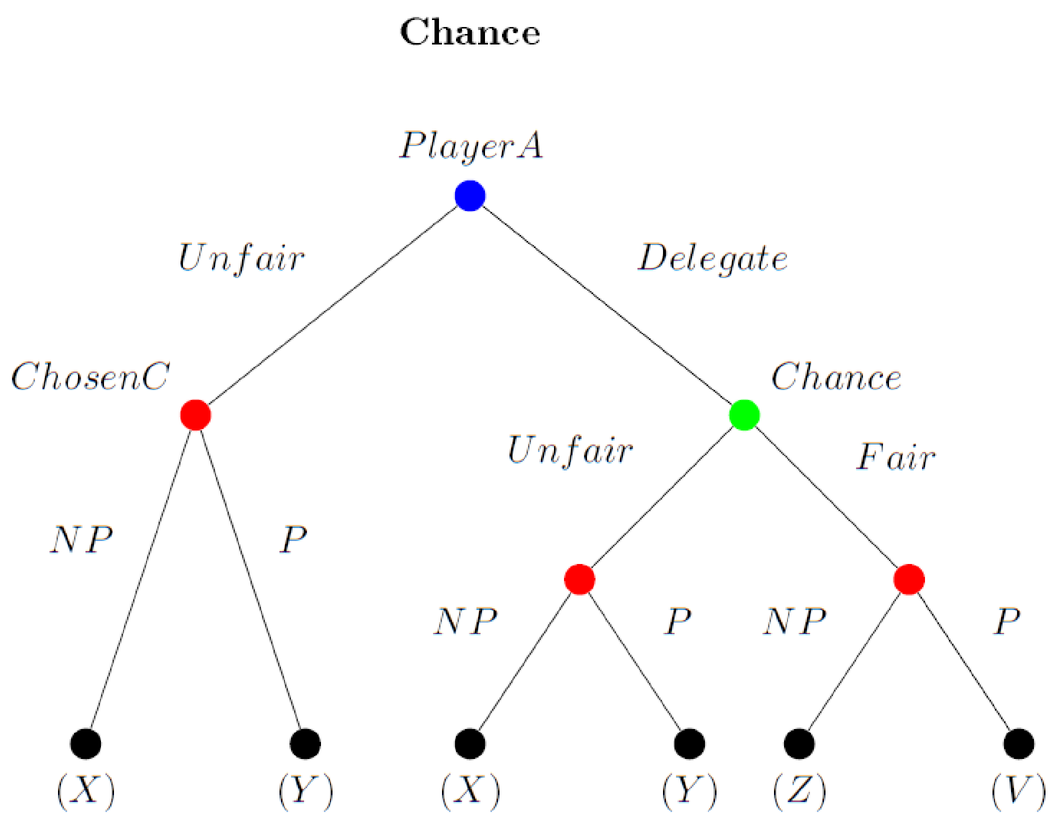


Figure 2: <u>Game Tree *Chance* treatment</u>

Our inter-treatment comparison investigates the second research question of this paper: “Do unkinder intentions by the dictator cause more displaced aggression towards innocent bystanders?” The within-subject comparison of the response to unfairness in *Chance* tests our third research question: “Do bad intentions cause more displaced aggression than bad luck?”

We were interested in seeing what proportion of subjects would have liked to punish if they could directly retaliate against the unfair dictator. Therefore, after stating their punishment decisions for all the possible payoff-relevant scenarios, Player Cs were presented with parallel hypothetical scenarios where Player A could also be punished. They were then asked to state their punishment preferences for these hypothetical scenarios.

After the two payoff-relevant stages of the game, all subjects in both of the treatments were asked to estimate what proportion of Player As in their opinion directly chose the unfair distribution (vs. the fair distribution in *Baseline* and vs. delegation to chance in *Chance*).

In Appendix D, we provide the translated instructions and control questions for the players in the role of Player C in the *Chance* treatment. Instructions in German for the other treatments and/or roles are available upon request.

2.2. Subject Pools

Our fourth research question considers sub-cultural differences in reaction to potential unfairness. A particular subculture comes to mind when one thinks of potentially frustrating situations. Soldiers must partake in conflict situations because of their professional demands. Moreover, the military system is ridden by a strong hierarchy. Soldiers usually cannot talk back, so frustration might stay “bottled up”. Therefore, we are particularly interested in exploring how soldiers react to unfair situations where the potential wrongdoer cannot be retaliated against. Thus, we conducted experiments at a university which is affiliated with the military. This subject pool will be referred to as *Soldiers* from now on.

As a control for this special subject pool, we conducted the same treatments (*Baseline* and *Chance*) with students at a regular university. This subject pool from now on will be referred to as *Students*.

Considering the professional challenges of individuals serving in the military, it is a relevant question whether soldiers are more or less likely than regular students to engage in displaced aggression. On the one hand, discipline seems to be a core value in the military sub-culture. A more evolved sense of discipline might translate into a less emotional approach to unfairness, and therefore less displaced aggression. On the other hand, particularly dominant and aggression-prone individuals possibly self-select into the military profession.

As we compared our *Soldiers* and *Students* subject pools, we felt concerned that besides the difference in career self-selection, which is the variation of interest for our research question, these two subject pools might differ in two more ways:

1) Students at a regular university might have a lower interconnectedness level in comparison to the strong social ties established among the *Soldiers*. Soldiers spend during their student days a lot of time together outside of the classroom, including military training and leisure at the shared university dormitories. Michailova and Bühren (2015) argue that due to the trained comradeship of soldiers, students at a university of the armed forces belong to a group with stronger social ties than a group of students at any other regular university. Similarly, Wiens et al. (2015) find evidence that cooperation and social responsibility of soldiers are higher than that of other students.

2) Experimental sessions at the regular university had a much lower proportion of male subjects (51%) than the lab sessions at the military university, where the majority of experiment participants (92%) were male. Although we can certainly control for gender effects when considering individual subjects' decisions, we were concerned that the gender composition of the co-players in the session that one can observe might also affect one's decisions in the experiment.

To counteract these concerns, we complemented our study with another pool of subjects who also have strong social ties and are mostly male. We recruited teams of Ultimate Frisbee players at a big tournament. Most of the Frisbee players are students of different disciplines. This subject pool will be mentioned as ***Frisbees*** from now on.

Ultimate Frisbee is a very special sport since it is played without any referee. After each Frisbee match, teams discuss fairness aspects of the specific game and rate their opponent team with spirit points depending on fairness. Frisbee players build a strong community – during tournaments, which are typically played over the weekends, all the players camp together.[7]

Moreover, due to the natural composition of the teams, we were able to conduct mostly male sessions (94% male), without calling any attention of the subjects to the gender composition of the sessions.

The subject pool of the Frisbee players could not be recruited to come to the lab. These players completed a pen and paper version of the game directly at the Frisbee tournament location. We only conducted the Chance treatment with the Frisbee players.[8]

[7] Please see a short video (in German) about this sport and the championships under the following link: https://www.frankenfernsehen.tv/mediathek/video/deutsche-meisterschaft-im-ultimate-frisbee/ (Franken Fernsehen, 2013).

[8] We conducted two more small (12 persons each) pen-and-paper sessions to control for the potential differences between the computerized and pen-and-paper versions of the *Chance* treatment: No significant differences can be observed.

3. Theory

In the following, we describe two economic behavioral models that compete to predict the behavior of subjects in our experiment – the frustration model by Battigalli et al. (2019) and the inequity-aversion model by Fehr and Schmidt (1999). Pfister and Böhm (2012) distinguish between anger and inequity-aversion in their three-person ultimatum game and find that rejections of responders are driven by anger, and not by inequity-aversion.

The behavioral model of frustration (Battigalli et al., 2019) explicitly considers an example of displaced aggression and therefore is fitting for our research questions. This model from now on will be referred to as either the B-D-S model or the Frustration model.

To calculate utility using this model, one must first calculate the frustration level caused by a specific scenario. The frustration level depends on the actions of others and one's expectation for those actions. Incidents can either be frustrating (i.e. have a positive frustration level) or non-frustrating (i.e. a frustration level equal to zero). Intuitively frustration is the difference between what one initially expected and what one can get, given the choices made by other(s) and one's alternatives in this situation (see also Aina et al., 2020, and Persson, 2018).

Given the potentially frustration-provoking design of our game, where the punishable players (B and other C) cannot be blamed for the provoked frustration, utility according to the B-D-S model depends on one's payoff in a particular scenario, the payoff of all the other groupmates in that particular scenario, the frustration level generated by this scenario, and one's own internal irritability level, which is part of one's personality. Given the payoff structure of the game, we will calculate what the parameter for the irritability must be for Player C to choose punishment in a particular scenario. Please see Appendix A for details. In *Baseline*, Player C should choose to punish Player B by one point if his irritability level is high enough: $\theta_C > {}^{1}/_{5F_C(Unfair)}$, where F_c is the frustration level.

Alternatively, the *payoff-driven theory of fairness* by Fehr and Schmidt (1999) is a candidate for evaluating our research questions. According to this behavioral model, agents care about their

- An all-female Frisbee team from the championship
- A mixed-gender mixed-subject students session recruited at beach volleyball facilities

Neither of these control pen-and-paper sessions is significantly different in subjects' choices from *Frisbees* or *Students*. Therefore, we could conclude that our pen-and-paper sessions with *Frisbees* are design-wise no different from our computerized sessions with the *Students* and *Soldiers*.

payoff and their relative standing in comparison to the other players in their group. This model in our paper from now on will be referred to as either the F-S model or the Inequity-aversion model. According to the predictions of this model, Player C should choose to punish Player B by one point if envy (α) is sufficiently high: $\alpha_C > 3/2$. Please see Appendix B for details.

For the remainder of this section, all the punishment decisions will first be looked at through the lens of the B-D-S frustration model, and thereafter the inequity-aversion model by Fehr and Schmidt (F-S). Moreover, when applicable, additional *intuitive* predictions or predictions based on social psychology literature will also be discussed.

3.1. Inter-Treatment Comparison

When comparing the two (main) treatments, the dictator's direct choice for an unfair distribution signals worse intentions in *Chance* than in *Baseline* because opportunity costs of not choosing unfair are higher in *Baseline* than in *Chance*.[9] Whereas in *Baseline*, the alternative to the unfair distribution is directly a fair distribution, the alternative in *Chance* still results in an unfair distribution in 50%, which could still benefit Player A and be disadvantageous for Player Cs. Directly choosing the unfair distribution in *Chance* does not even give the fair distribution a chance. Therefore, possibly Player Cs are more frustrated by a direct unfair choice in *Chance* compared to *Baseline*. By this logic, they might engage in more displaced aggression in *Chance* than in *Baseline*.

1) B-D-S predictions.

The B-D-S (frustration) model predicts the opposite: Less punishment in *Chance* than in *Baseline*. The difference in expected payoffs for Player Cs between the treatments is the cause. In *Baseline*, when Player A is fair, the payoff is 25 points for each of the groupmates. In *Chance*, if Player A decides to delegate the distributive decision to chance (computer or coin), the expected payoff for Player Cs is only 15 points ($= \frac{5+25}{2}$). Hence for a Player C, the best-case scenario (when the dictator abstains from the unfair distribution) in terms of expected payoff is worse in *Chance* than in *Baseline*.

[9] *Baseline*: 45 – 25 = 20 vs. *Chance*: 45 - $\frac{45+25}{2}$ = 10.

The B-D-S model assumes that subjects are more frustrated by bigger potential losses. Therefore, according to B-D-S, Player Cs are expected to engage in more displaced aggression in *Baseline* than in *Chance*.

2) F-S predictions.

Since the dictator's direct choice for unfairness results in the same distribution in both treatments, the F-S model in both treatments predicts the same punishment behavior on the part of Player C.

3.2. Within-Subject Comparison

Our *Chance* treatment allows for a within-subject comparison of punishment behavior. The two unfair distribution scenarios can either be caused by bad intentions caused by the dictator's choice, or bad luck resulting from random chance. If intentions matter, subjects should be more frustrated and therefore more aggressive when a disadvantaged position has been caused by the whim of their co-player rather than random chance.

1) B-D-S predictions.

The B-D-S model predicts approximately equal levels of displaced aggression when Player A is directly unfair or when Player A delegates her decision to chance, which results in an unfair distribution. Please see Appendix C for details.

2) F-S predictions.

According to the F-S model, it should not matter if the unfair distribution has been caused by A's bad intentions or bad luck randomly generated by either the computer or the coin flip (depending on the session). If subjects have inequity-averse preferences, these two scenarios of *Chance* treatment have the same payoff structure, and therefore must result in the same punishment behavior, which depends on the individual's envy parameter α. Subjects with a sufficiently high α (>1.5) should punish Player B in both scenarios. For these sufficiently-envious player Cs, this punishment pattern would bring their payoff closer to the payoff of their co-players and would thereby boost their utility.

3.3. Who Should Get Punished? (in unfair distributions)

Marcus-Newhall et al. (2000)[10] show that when frustrated, people are more likely to punish an innocent bystander, if the latter is somehow similar to the person who has wronged them. Our experimental design accommodates a renewed test of this theory: each group in our game has one punishable co-player that is similar to the dictator (Player B), and another punishable co-player that is similar to the punishing agent (the other Player C). By this logic, frustrated Player C should punish the Player B, but not the other Player C.

1) B-D-S predictions.

   Since utility according to the B-D-S model depends on the sum of co-players' payoffs, there should be no difference in Player C's utility when punishing by one point the passive Player B (who is "rich" like the dictator) or the other Player C (who is "poor"). The condition for one's irritability level is the same: $\theta_C > {}^{1}\!/_{5F_C(Unfair)}$.

2) F-S predictions.

   A Player C with F-S preferences should never punish the co-playing Player C. Since unfair distribution leaves the two Player Cs with equal endowments, the punishment of the other Player C could only cause disutility to the punishing Player C.

3.4. Should Subjects Punish in Fair Distributions?

There are two scenarios in which the fair distribution is possible: when Player A directly chooses the fair distribution in *Baseline*, or when in *Chance* Player A delegates her decision to chance, which generates a fair distribution.

1) B-D-S predictions.

   Frustration level equals zero, when distribution is fair.[11] The B-D-S model predicts no displaced aggression if one is not frustrated.[12]

2) F-S predictions.

---

[10] A meta-study based on 82 papers in social psychology. Interestingly, for many of these studies, displaced aggression is not the direct research focus.

[11] In *Baseline*: $F_C(Fair)_{Baseline} = \max\{0, 5\varepsilon + (1-\varepsilon)[25(1-r) + 24r] - max(25, 24)\} = \max\{0, -20\varepsilon - r - \varepsilon r\} = 0$.

In *Chance*: $F_C(Comp\_Fair)_{Chance} = \max\{0, -10 - 10\varepsilon - 0.5r + 0.5\varepsilon r\} = 0$

[12] Punishment is costly: $U_C(Fair, NotPunish) = 25 - 0$ vs. $U_C(Fair, Punish) = 24 - 0$.

In the case of fair distribution, all four co-players have an equal endowment. Punishment in this case can only create disutility. Therefore, according to the F-S model, no punishment should be observed in the case of fair distribution.

4. Results

4.1. "Direct Unfairness"

The proportion of unfair Player As is not significantly different between Baseline and Chance, neither for Students nor for Soldiers. When aggregating across those two types of treatments, unfairness rate of Soldiers is weakly significantly higher than the unfairness rate of Students (2-sided Fisher's exact test, p = 0.064), and is significantly higher than the unfairness rate of Frisbees (2-sided Fisher's exact test, p = 0.006).

Treatment / Subject Pool	**Baseline**	**Chance**
Students	58% (N=12)	62% (N=13)
Soldiers	83% (N=6)	100% (N=6)
Frisbees	-	38% (N=13)

Table 1: *Proportion of direct unfairness by subject pool and treatment*

The average direct unfairness rate for all observations in Table 1 is 62%. Average hypothetical "direct unfairness" rate stated by passive Player Bs is only slightly lower, at 56%.

Moreover, all three types of players (A, B, C) are good at predicting the direct unfairness rate for their subject pool, although *Frisbees* often overestimate the proportion of unfair dictators. The average estimate of direct unfairness equals 70%. (When isolating just the beliefs of Player Cs, the average estimate of dictators' direct unfairness is 68%).

Among the *Students* subject pool, which has an almost equal number of male and female subjects, female subjects are more generous than their male counterparts. 33% of all female *Students* directly choose the unfair distribution, compared to the 75% of all the male *Students* who do the same (2-sided Fisher's exact test, p = 0.087). Yet, both men and women have similar and close to reality beliefs regarding the proportion of unfair dictators (72% vs. 68%, Mann-Whitney test, p = 0.339).

4.2. The Proportion of Economically Rational Players

According to standard economic theory, subjects should not invest any of their payoffs into the punishment of their co-players if they are only interested in their payoff. Since punishment is costly, we define subjects as “economically rational" if they decide not to punish their co-players in any of the presented scenarios.

Treatment / Subject Pool	**Baseline**	**Chance**
Students	71% (N=24)	73% (N=26)
Soldiers	50% (N=12)	25% (N=12)
Frisbees	-	69% (N=26)

Table 2: *Proportion of economically rational players by subject pool and treatment*

Although most subjects are economically rational (altogether 63%), a considerable proportion of subjects (37%) engage in inappropriate punishment in at least one of the scenarios they are presented with. *Soldiers* seem more susceptible to inappropriate punishment than others (63.5% vs. 29% among *Students* and *Frisbees* pooled together, 2-sided Fisher’s exact test, p = 0.007).

Most of the following results are discussed in terms of binary decision to punish or not to punish. The main results are no different if we instead consider the total amount of points invested in punishment.

4.3. Punishment Behavior in Unfair Scenarios

	Inter-Treatment Comparison (Baseline vs. Chance)		Within-Subject Comparison (Chance vs. Chance)
Treatment	**Baseline**	**Chance**	**Chance**
Unfairness By…	Dictator	Dictator	Computer / Coin flip
Students	29% (N=24)	23% (N=26)	27% (N=26)
	p = 0.751		p = 1.000
Soldiers	42% (N=12)	75% (N=12)	50% (N=12)
	p = 0.214		p = 0.250
Frisbees	-	23% (N=26)	19% (N=26)
			p = 1.000

Note: 2nd lines for each subject pool provide 1) p-values of 2-sided Fisher's exact tests for the Inter-Treatment Comparisons, and 2) for Within-Subject Comparisons, the p-values of the McNemar's exact test (commonly used for within-subject observations).

Table 3: *Punishment Behavior in Unfair Scenarios*

4.3.1. Inter-Treatment Comparison

When comparing *Baseline* and *Chance*, differences in the punishment rate are not significantly different for neither *Students* nor *Soldiers* (2-sided Fisher's exact test p = 0.751 and p = 0.214, respectively; p = 0.635 when these two subject pools are combined). This is in line with the predictions of the F-S model.

4.3.2. Within-Subject Comparison

When aggregating across all the three subject pools in *Chance*, 18 out of the 64 Player Cs (28%) choose to punish their co-players in the scenario in which the unfair distribution has been caused by chance. In comparison, 21 out of the 64 subjects C (33%) choose to punish their co-player(s) when the dictator in *Chance* directly chose the unfair distribution. Hence, for most subjects, the source of inequality is not important. This empirical result is in line with both F-S and B-D-S models' predictions.

4.3.3. Who Gets Punished?

Among all the available unfair distribution cases, Player Cs chose to punish in 51 cases (out of the total possible 164 cases). In 63% of these cases, the only target of punishment was Player B. These observations are in line with the findings of the Marcus-Newhall et al. (2000) meta-study since the "rich" passive Player B resembles the "rich" wrongdoing Player A.

Subject pool	**Treatment**	**Unfairness by…**		**B only**	**C only**	**B & C**	**Punishments N**
Students	Baseline	Dictator		57%	0%	43%	7
	Chance	Dictator		50%	0%	50%	6
		Computer		71%	0%	29%	7
Aggregate of Students				**60%**	**0%**	**40%**	**20**
Soldiers	Baseline	Dictator		40%	0%	60%	5
	Chance	Dictator		56%	0%	44%	9
		Computer		67%	0%	33%	6
Aggregate of Soldiers				**55%**	**0%**	**45%**	**20**
Frisbees	Chance	Dictator		83%	0%	17%	6
		Coin flip		80%	0%	20%	5
Aggregate of Frisbees				**82%**	**0%**	**18%**	**11**
Aggregate of all observations				**63%**	**0%**	**37%**	**51**

Table 4: *Who gets punished? (when punishment is executed in the unfair scenarios)*

Although no subject in the role of Player C punishes just the co-playing Player C, of all the punishments exercised as a response to unfairness, 37% involve not only the passive Player B but also the co-playing Player C. The latter is not only dissimilar to the potentially wrong-doing Player A but is the same type of player as the punishing agent himself.

The B-D-S model can partially explain the observed punishment preferences of Player Cs. According to the predictions of the B-D-S model, it should not matter for the subjects which of their "innocent" co-players they are punishing. Both players could be "in the wrong place at the wrong time" (frustration level is the same). However, by this same logic, the B-D-S model fails to explain the absence of Player Cs who only punish the co-playing Player C, but not Player B.

The F-S model of inequity-aversion, in its original form, cannot explain the behavior of Player Cs who punish the co-playing Player C. In all the scenarios of unfair distribution, both Player Cs have an endowment of 5 points each. Since they have an equal payoff to begin with, the punishment would create inequality, which in turn could only decrease the punishing party's utility.

However, it is possible that relatively many subjects choose to punish Player C along with Player B because they do not wish to be the worst-off person in their group.

A modified version of the F-S model that allows for a negative β parameter: "status-seeking" preferences could explain why the chosen Player C might punish a co-player that has the same payoff he has.[13]

4.3.4. Punishment Behavior in Fair scenarios

Treatment / Subject Pool	**Baseline** (A's Decision)	**Chance** (Lucky Chance)
Students	12.5% (N=24)	19% (N=26)
Soldiers	17% (N=12)	25% (N=12)
Frisbees	-	15% (N=26)

Table 5: *Punishment in fair scenarios*

Punishment rates in the fair scenarios are not significantly different across any of the possible comparisons. In summary, 17 out of our 100 Player Cs chose to punish their co-player(s) in one of the fair scenarios, caused either by Player A's direct fairness or Player A's delegation combined with fair distribution caused by chance. The observed behavior is somewhat similar to the status-seeking behavior in the famous joy of destruction game by Abbink and Herrmann (2011), with confirmed validity also in the field setting (Prediger et al., 2014).

Interestingly, among cases of punishment, on average more points are invested into punishment when the distribution is fair rather than unfair: 3.7 points vs. 2.9 points, respectively;[14] which translates into 18.5 points vs. 14.5 points of damage to co-players, respectively. Similarly, when punishing their co-players in the fair scenarios rather than unfair scenarios, subjects punish their co-players more often by the maximum amount of points possible (5 points): 35% (6 out of 17) vs. 23.5% (12 out of 51) of the time, respectively (but not significantly more according to a two-sided Fisher's exact test, p=0.357). These punishment trends could be due to the income effect in the fair

[13] The "status-seeking" β must then be inversely correlated to the subject's α (i.e. envy) parameter. Subjects that have a high α parameter, must have an especially low (in the negative domain) β parameter to also punish the other player C: $\beta < -\frac{\alpha+0.75}{2}$.

[14] Subjects who punish in both fair and unfair scenarios invest significantly more points into punishment in the fair scenarios: Wilcoxon-signed rank test, p=0.007.

distribution scenarios, meaning that Player Cs have more resources to spend on punishment. However, they still cannot invest more than 5 points in the punishment of co-players.

The B-D-S frustration model cannot explain any of the punishment we see in the fair distribution scenarios. The level of frustration in the fair distribution scenarios always equals zero. When frustration equals zero, no utility can be gained from punishment. Therefore, no subject should engage in punishment in any of the fair distribution scenarios.

In our post-experimental questionnaire, when asked about their general strategies in the game (an open-ended question), quite a few subjects expressed a desire for status. Perhaps most eloquently, one subject wrote: “I do not want that everybody has the same amount of money. It’s better when other participants have less than me.”[15]

In the fair distribution scenarios, the two non-impune co-players, Player B and the other (non-chosen) Player C, have the same payoff as the punishing party. Therefore, here, a Player C who wishes to be better off than others should punish both Player B and Player C. This explains why only 3 Player Cs (out of the 17 "fair"-punishers) choose to punish only Player B and no one punishes just the other Player C.

If we look at this phenomenon with the F-S model in mind, this behavioral trend points towards a negative β factor, which is not foreseen by the original F-S model. A modified version of the F-S model that allows a negative β parameter could explain such status-seeking behavior. Moreover, the same parameter restrictions would apply as in the previous sub-section 4.2 (on punishment of co-playing Player C in unfair distributions): $\beta < -\frac{\alpha+0.75}{2}$.

Hence, individuals that have the corresponding α and β parameters should engage in punishment in both directions: (also) punishing Player C in the unfair distribution scenarios and punishing both co-players in the fair distribution scenarios. However, these within-subject predictions are not always consistent with our empirical evidence.

		Punish (also) C in Direct Unfairness scenario	
		Yes	*No*
Punish in Fair Scenarios	*Yes*	7	10
	No	7	76

[15] Translated from German: “Ich möchte nicht, dass alle exakt dieselbe Punktzahl haben. Besser, wenn andere Teilnehmer weniger als ich haben”

Table 6: *Within-subject punishment behavior in fair and unfair scenarios*

17 (= 7 + 10) subjects choose to punish in one of those status-improving scenarios, but not in the other. Additionally, there are 3 subjects (out of all 100 Cs) that punish co-players only in case of fair distribution.

4.4. Discussion of Subject Pool Differences

There are no differences in any of the punishment rate comparisons between *Students* and *Frisbees*. Hence, differences in the interconnectedness level among subjects or gender composition of the sessions do not play a role. For the discussion in this sub-section, we pool together these two subject pools as *Non-Soldiers*.

When aggregating across the two treatments, *Soldiers* punish similarly to *Non-Soldiers* in the fair distribution scenarios (2-sided Fisher's exact test, p = 0.547), but punish significantly more than *Non-Soldiers* in the direct unfairness scenario (2-sided Fisher's exact test, p = 0.005). In the unfair distribution scenario, generated by a computer or a coin flip, punishment behavior between *Soldiers* and *Non-Soldiers* is weakly significantly different (2-sided Fisher's exact test, p = 0.080).

Self-selection into a military career seems to be an important determinant of action when looking at inappropriate punishment. Our data demonstrate that in the unfair distribution scenarios *Soldiers* are more likely to engage in punishment of their co-players than *Non-Soldiers*. If we assume that subjects have F-S preferences, a higher proportion of *Soldiers* (than *Students* or *Frisbees*) must have an $\alpha_C > 3/2$.

	Model (1)	Model (2)	Model (3)	Model (4)
Soldier	0.333***	0.342***	0.370***	0.365***
Chance	-	0.021	0.021	0.021
Male	-	-	-0.111	-0.115
Impatience Measure	-	-	-	0.029
N_C	100	100	100	99

Note: *** corresponds to a 1% significance level. N_C is the number of subjects C.

Table 7: *Marginal effects on punishing of probit regressions for the direct unfairness scenario*

If we assume that subjects have B-D-S preferences, given that *Students* and *Soldiers* on average expect similar direct unfairness rates from the dictator (among Player Cs: 71% vs. 77%, respectively; Mann-Whitney test, p = 0.293), the frustration level $F_C(Unfair)$ experienced by *Soldiers* and *Non-Soldiers* is not much different. Therefore, for *Soldiers* to engage in more displaced aggression in the unfair scenarios, they must have on average a higher irritability parameter θ_C than *Non-Soldiers*. To punish a co-player by one point, *Soldiers* are more likely to fulfill the $\theta_C > 1/5F_C$ condition.

5. Discussion and Conclusion

Model	Inter-Treatment Comparison		Within-Subject Comparison		Punishment of Pl. C (any scenario)		Punishment in Fair scenarios	
B-D-S	Baseline > Chance	✗	Baseline ≈ Chance	✓	Yes	✓	No	✗
F-S	Baseline = Chance	✓	Baseline = Chance	✓	No	✗	No	✗
F-S + Status	Same as F-S	✓	Same as F-S	✓	Yes	✓	Yes	✓

Table 8: *Summary of Theory Comparisons - Predictions and Results*

We see that a large proportion of our subjects (63% in the main treatments) show economically rational behavior. For the remaining 37%, a variation of the F-S model, the F-S + Status, seems to best explain our experimental evidence. Punishing subjects are inequity-averse in all the scenarios in which they can be in a disadvantaged position. The punishment that aims to close the payoff gap between themselves and the punishable "rich" Player B is not considerably different between the main two treatments or between scenarios within *Chance*.

Yet, a sub-group of those “disadvantage-averse” subjects choose to create inequality when they are not at a disadvantage. Punishments in the fair scenarios and punishment of the other Player C in the unfair scenarios can be explained neither by standard theory nor by either of the theories we considered. It seems unreasonable to invest resources into the punishment of "equals" unless one wants to be better off than them.

Possibly, the unfair power structure of our game activated among subject an updated social norm of unfairness. Maybe, the possibility of being treated unfairly, especially without adequate options for retribution, and correct realization that such treatment is very likely, awakened subject's not even selfish but spiteful side.

Summary of results:

- 1st research question: A sizable proportion of subjects pay for the "sins" of the unfair impune dictator.
- 2nd research question: Dictators' *unkindness* level does not seem to play a role.
- 3rd research question: Punishers seem to be equally reactive to unfavorable circumstances and unfair treatment directed at them: *bad luck* vs. bad intentions, respectively.
- 4th research question: Association with the military sub-culture seems to affect punishment behavior in unfair scenarios.

The motive for inappropriate punishment as a way of "letting off steam" has been overlooked in the previous literature. It would be interesting to study the role of displaced aggression in future experiments. To differentiate between inequity-averse and "emotional outbursts" type of displaced aggression, it might be insightful to conduct an experiment where a passive player's exact endowment is unknown to the potential punisher. As an example we could suggest a two-stage game in which the first stage is a simple two-person dictator game, and in the second stage the recipient in the dictator game can pay to punish a passive player whose endowment is randomly generated and is unknown to the other players.

Declaration of Competing Interests

The authors declare no conflict of interests.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abbink, K., Herrmann, B., 2011. The Moral Costs of Nastiness. Econ. Inq. 49 (2), 631–633. 10.1111/j.1465-7295.2010.00309.x.

Aina, C., Battigalli, P., Gamba, A., 2020. Frustration and anger in the Ultimatum Game: An experiment. Games Econ. Behav. 122, 150–167. https://doi.org/10.1016/j.geb.2020.04.006.

Bartling, B., Fischbacher, U., 2012. Shifting the Blame: On Delegation and Responsibility. Rev. Econ. Stud. 79 (1), 67–87. https://doi.org/10.1093/restud/rdr023.

Battigalli, P., Dufwenberg, M., Smith, A., 2019. Frustration, aggression, and anger in leader-follower games. Games Econ. Behav. 117, 15–39. https://doi.org/10.1016/j.geb.2019.06.001.

Bushman, B.J., Baumeister, R.F., Stack, A.D., 1999. Catharsis, aggression, and persuasive influence: Self-fulfilling or self-defeating prophecies? J. Pers. Soc. Psychol. 76 (3), 367–376. https://doi.org/10.1037/0022-3514.76.3.367.

Card, D., Dahl, G.B., 2011. Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. Q. J. Econ. 126 (1), 103–143. https://doi.org/10.1093/qje/qjr001.

Dickinson, D.L., Masclet, D., 2015. Emotion venting and punishment in public good experiments. J. Public Econ. 122, 55–67. https://doi.org/10.1016/j.jpubeco.2014.10.008.

Dollard, J., Miller, N.E., Doob, L.W., Mowrer, O.H., Sears, R.R., 1939. Frustration and aggression. Yale University Press, New Haven, CT, US.

Drouvelis, M., Grosskopf, B., 2016. The effects of induced emotions on pro-social behaviour. J. Public Econ. 134, 1–8. https://doi.org/10.1016/j.jpubeco.2015.12.012.

Fehr, E., Schmidt, K.M., 1999. A Theory of Fairness, Competition, and Cooperation. Q. J. Econ. 114 (3), 817–868. https://doi.org/10.1162/003355399556151.

Fehr, E., Gächter, S., 2000. Cooperation and Punishment in Public Goods Experiments. Am. Econ. Rev. 90 (4), 980–994. https://doi.org/10.1257/aer.90.4.980.

Fessler, D.M.T., 2006. The Male Flash of Anger: Violent Response to Transgression as an Example of the Intersection of Evolved Psychology and Culture, in: Barkow, J.H. (Ed.), Missing the Revolution. Oxford University Press, pp. 101–118.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178. https://doi.org/10.1007/s10683-006-9159-4.

Franken Fernsehen, 2013. Deutsche Meisterschaft im Ultimate Frisbee. https://www.frankenfernsehen.tv/mediathek/video/deutsche-meisterschaft-im-ultimate-frisbee/. Accessed 8 July 2020.

Gurdal, M.Y., Miller, J.B., Rustichini, A., 2013. Why Blame? J. Political Econ. 121 (6), 1205–1247. https://doi.org/10.1086/674409.

Kaiser, J., Pedersen, K., Koch, A., 2018. Do Economists Punish Less? Games 9 (4), 75. https://doi.org/10.3390/g9040075.

Marcus-Newhall, A., Pedersen, W.C., Carlson, M., Miller, N., 2000. Displaced aggression is alive and well: A meta-analytic review. J. Pers. Soc. Psychol. 78 (4), 670–689. https://doi.org/10.1037/0022-3514.78.4.670.

Michailova, J., Bühren, C., 2015. Money Priming and Social Behavior of Natural Groups in Simple Bargaining and Dilemma Experiments. MAGKS Joint Discussion Paper Series in Economics (30-2015).

Mintz, A., Redd, S.B., Vedlitz, A., 2006. Can We Generalize from Student Experiments to the Real World in Political Science, Military Affairs, and International Relations? J. Confl. Resol. 50 (5), 757–776. https://doi.org/10.1177/0022002706291052.

Mocan, N., 2013. Vengeance. Rev. Econ. Stat. 95 (3), 969–982. https://doi.org/10.1162/REST_a_00325.
Persson, E., 2018. Testing the impact of frustration and anger when responsibility is low. J. Econ. Behav. Organ. 145, 435–448. https://doi.org/10.1016/j.jebo.2017.12.001.
Prediger, S., Vollan, B., Herrmann, B., 2014. Resource scarcity and antisocial behavior. J. Public Econ. 119, 1–9. https://doi.org/10.1016/j.jpubeco.2014.07.007.
Pfister, H.-R., Böhm, G., 2012. Responder Feelings in a Three-Player Three-Option Ultimatum Game: Affective Determinants of Rejection Behavior. Games 3 (1), 1–29. https://doi.org/10.3390/g3010001.
Sánchez-Jankowski, M., 1991. Islands in the street: Gangs and American urban society. University of California Press, Berkeley.
Simon, H.A., 1955. A Behavioral Model of Rational Choice. Q. J. Econ. 69 (1), 99. https://doi.org/10.2307/1884852.
Wiens, M., Johannemann, K., Morasch, K., Hofmann, M., 2015. Offizier und Gentleman? Eine experimentelle Untersuchung berufsbezogener Normen am Beispiel des Offiziers. Volkswirtschaftliche Diskussionsbeiträge 2015,2, Neubiberg. http://hdl.handle.net/10419/125854.