FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Purely Procedural Preferences in Auctions, with a comment to: Equality of opportunity and the acceptability of outcome inequality

Nadine Chlaß

JENA ECONOMICS RESEARCH PAPERS · # 2026 – 009

The JENA ECONOMICS RESEARCH PAPERS
is a publication of the Friedrich Schiller University Jena, Germany (www.jenecon.de).

Purely Procedural Preferences in Auctions, with a comment to: Equality of opportunity and the acceptability of outcome inequality

Nadine Chlaß

Department of Economics
Faculty of Economics and Business Administration
Friedrich-Schiller-University Jena
Carl-Zeiss-Str. 3, 07743 Jena, Germany.
Email: nadine.chlass@uni-jena.de
Phone: 0049 (0)3641 943243

Abstract

Sugden and Wang (2020) are among the first to model intrinsic concerns for equal opportunities. Their study reports that after being subjected to unequal opportunities, people experience anger, resentment, and greed and all parties involved engage in socially wasteful destruction of resources. In this critique, I point out errors which compromise these findings. The authors study two players in a card game where one player is dealt more replacement cards (unequal opportunities); the highest card wins. These replacement cards must not change the expected outcome of the game if they are to have intrinsic value only. If this is true, however, replacement cards cannot induce distinct expected outcomes, which they must do if they are to constitute opportunities by the authors' own definition. In short, if replacements *do* yield distinct expected outcomes, they *must* inevitably also have instrumental (material) value which confounds the treatment effect. The authors then elicit people's emotional response to these unequal replacement options in a vendetta game, where players can take and partly destroy each others' payoff. What remains unnoticed is that the vendetta game itself introduces unequal opportunities *along every potential path of play*. I append an earlier study which uses a different approach to identify the behavioral consequences of people's intrinsic concerns for unequal opportunities: In a game which grants one party infinite opportunity and only little to her opponent, players' departures from narrow rational self-interest are explained by two redundant instruments for equal effective opportunities (Chlaß et al., 2019, Chlaß and Riener, 2015/2025). The party with greater opportunity compensates her opponent; the opponent requires to be compensated or else sets all parties' payoff to zero.

JEL classification: C91, D02, D44, D63, D82, D91

Keywords: equality of opportunity, decision rights, autonomy, power, common value auction, market for lemons, information asymmetry

1 Critique

Trautmann (2023) lists two papers which model players that value not only their own, but also their opponents' freedom of choice intrinsically – Chlaß et al. (2019) and Sugden and Wang (2020). I report errors in the latter which compromise the main findings therein. Sugden and Wang (2020) intend to show that people intrinsically care for (un)equal strategic opportunities beyond the outcomes and kindness which these opportunities entail; they report that all parties involved respond with destructive behaviour to such inequality. Their paper introduces a game which treats two parties asymmetrically to induce unequal opportunities: In a card game, one party is dealt more replacement cards than her opponent; the highest card wins the game. The unequal number of cards (opportunities) represents the treatment intervention. Subsequently, both parties play a vendetta game to measure their emotions – the response variable. In this vendetta game, the loser of the card game begins; she may take payoff away from her opponent in units of three. By doing so, she destroys two of these three units. Next, her opponent may do the same, until either in four subsequent decision nodes, no party took anything, or no more three units may be taken from either party.

What remains unnoticed is that the vendetta game assigns systematically more opportunities to the loser $\mathscr{L}$ of the card game than to its winner $\mathscr{W}$ and therefore introduces *systematic inequality of opportunity*. In short, the measurement tool by which the authors elicit players' emotions after unequal opportunities, distorts the response variable. To demonstrate this, Appendix A first shows the vendetta game with a simplified stopping rule, numbering all potential courses of play across its four subtrees, that is, all possible ways in which the game may unfold. For each path, the Appendix counts the number of options each player has when called upon to play. Along *all but six* paths of play, the loser of the card game who begins the vendetta game, has *more* options to choose from than her opponent. For the *remaining six* paths of play (path IDs *2-03, 2-04, 2-05, 3-02, 3-03, 3-04*), the loser of the card game also has more options *up to her final decision node*. Only here does she have fewer options than her opponent would subsequently have. Note, however, that the loser can at this very decision node, always choose such as to leave her opponent without any future choice at all. She can force her opponent onto a path where again, she always has greater freedom of choice. Hence, $\mathscr{L}$ can unilaterally prevent this last disadvantageous difference from ever materializing. In the actual experiment, $\mathscr{L}$'s advantage in opportunities was always made clear from one period to the next: A player's options were intuitively visualized on screen as shown in Online appendix A – the units which could be taken from the opponent in a given period; the units a player would add to her own payoff with each option (i.e. after 2/3 of the bounty had been subtracted); and the units in consequence which her opponent would be able to take in the next period. It does therefore not even matter how far a player reasons through the game. In most cases, $\mathscr{L}$ sees that she has more opportunities than her opponent, $\mathscr{W}$, and where not, sees how to opt into having more. Naturally, this inequality in how the rules of the game treat both players per path of play also translates into the number of *strategies* of the extensive game form if one wants to count opportunities this way. Table A8 in Appendix I on p. 37 counts the number of options along each potential path of play in the vendetta game without the simplified stopping rule. The issue remains.[1]

[1] One could argue that not all of these options imply different outcomes (a critical ingredient to the definition of an opportunity, as we will see next) and do therefore not all constitute opportunities. But

If we now recall that being dealt more replacement cards implies a higher chance of winning; that the prize for winning is the role $\mathscr{W}$ in the vendetta game; that $\mathscr{W}$ is endowed with nine lottery tickets whereas the loser receives only three; that these tickets are the units to be taken in the vendetta game; then we see that *both* parties have good reason to be upset, as the authors are surprised to observe. The loser may be upset for distributional reasons – for being more likely to lose the card game and for receiving only three lottery tickets which, at the end of the experiment, entail a monetary prize. The winner will be upset because, after all, she was more likely than the loser to 'win' her current disadvantage in opportunities in the vendetta game.

Measurement of opportunities. Let us now look at the treatment intervention. A core idea of the study is to design a game in which outcomes depend (predominantly) on luck such that the procedural asymmetry does not also cause ex-ante unequal payoffs. If it did, the asymmetry in opportunities would not only have the desired '*non-instrumental*' (p. 9) or *intrinsic* value, but also *instrumental* value, and all procedural preference models to date, whether outcome-based, intention-based, or purely procedural, would apply. Indeed, the authors advertise that this instrumental value is imperceptibly small (p. 10). This does, however, have unexpected consequences for the treatment intervention, that is, for the number of replacement cards being dealt. If the latter *do not entail different expected outcomes* since they may all with equal chance lead to one outcome rather than another, they do not induce distinct outcomes and do not constitute opportunities. The entire paper does, however, assume that they *do* both (p. 5). In short: *either* there is no treatment intervention at all, *or* the experiment studies players' emotional reactions to unequal expected payoffs.[2] One cannot help but think that the authors abandon the mathematical precision of their concept to the detriment of the study, for, as they put it, the sake of 'obviousness' [p. 7] and intuition.

A continuous treatment variable. So far, I have argued that, by construction, either there is no treatment intervention, or the emotional reaction must be confounded. Looking closely, there appears to be yet another problem. In the card game, each player is dealt a card and then decides to replace. The number of times she can replace is used to define her opportunities. However, the moment the player considers the first replacement, the value of this 'opportunity' hinges upon the value of her hand. For the player with several replacement cards, the value of each replacement is therefore contingent on the value of the previous one (her current hand). The treatment – i.e. the difference in opportunities – therefore, *differs* for each pair of players and game and is, in fact, a continuous random variable which the authors treat as binary. Certainly, the opponent's current card is unknown to the player, and therefore, whether or not a replacement card

since $\mathscr{L}$ always has more choice earlier on, she governs whether or not she opens subtrees in which options lead to identical outcomes; by directing the path of play, she governs whether an option at a given decision node induces a distinct outcome or not and whether therefore, that option is an opportunity.

[2] We can stress-test this objection a little further by arguing that, once a card is turned and the number it features known, the *turned* card may indeed induce an outcome different from a next card being turned. However, even this viewpoint does not arrange the matter, because the card game does not have continuous outcomes. Only the highest card wins a prize, all other cards yield the same payoff. Hence, even a just turned card will for many cases induce the same outcome for all players as yet another turned card does. Since sometimes a newly turned card may indeed induce a distinct outcome, there is a heterogeneous treatment intervention for each pair of players and each game. The player who can only replace a single card, may have more opportunity than her opponent with three (if, with each turned replacement card, the opponent still has the same payoff and does not hold the highest number). A similar argument can be made ex-ante, before a replacement card is turned, as I discuss next.

constitutes an opportunity is unknown to her, too. But the player's own current card is known to herself, and the expected difference in opportunities between her own known card, and the conditional expected values of the opponent's three cards, differs within each pair and within each game. Turning the authors' own example (p. 9), a player who is being dealt a high card and has (only) one option to replace, notices that a replacement card need not be an opportunity at all and that the opponent's three replacements need not, given the player's hand, change the expected outcome of the game. Even therefore, if there is a (confounded) treatment, the latter is not homogeneous.

Indication that there is, on average, no treatment intervention in the direction intended. Sugden and Wang find no evidence (p. 14) that players compensate these 'Unfair Rules' by adopting different strategies and becoming more altruistic (the advantaged party: leaving her additional replacement cards unused), a phenomenon which is observed in (Chlaß, 2010, Chlaß et al., 2019, Chlaß and Riener, 2015/2025, Chlaß and Miettinen, 2026). In the vendetta game, however, substantial altruism occurs in unexpected places: 32.5% of the losers – who hold the advantage in opportunity in the vendetta game – choose not to induce vendettas (p. 15) whereas 'winners are not inhibited to take' (p. 19). The authors themselves mention early on that, the more asymmetric players' initial positions in the vendetta game, the more vendettas are typically observed in the literature, too (p. 8). Once we appreciate that a player's initial position defines her opponent's set of opportunities, these findings are consistent with our result that players compensate asymmetric opportunities behaviourally.[3] The absence of such behavioural phenomena in the card game with 'Unfair Rules', while observed in our own studies *and* by the authors themselves in the vendetta game, supports the idea that the different number of cards did not induce unequal opportunities. Indeed, the authors themselves conjecture that '*...players did not recognise the strategic significance of the number of opportunities available to their opponents...*' (p. 14).

Relation to our own concept of equal effective opportunities. Finally, the authors commit errors regarding our concept of unequal decision rights, that is, unequal cardinalities of players' sets of opportunities. Chlaß (2010), Chlaß et al. (2019), and Sugden and Wang (2020) study the same hypotheses, i.e. whether a party disadvantaged in opportunities becomes less content with a given payoff even when these opportunities only have intrinsic value, and whether a party advantaged by such a procedure becomes more altruistic. The studies pursue the same idea and use similar language, i.e. 'rules treating players asymmetrically' when 'players have not chosen these rules' (Chlaß, 2010, p.1). Let me therefore briefly look at our own concept. A recent application to commitment games is found in (Chlaß and Miettinen, 2026). Chlaß et al. (2019) recursively construct a 'set of effective opportunities', starting with the option the player prefers least, then introduce her second least preferred, the third least preferred, and so forth, always provided the option to be added can be preferred over all other options already in the set, see (Jones and Sugden, 1982). For a narrowly self-interested decision maker, the set of effective opportunities will include all options which yield the decision maker distinct

[3] The disadvantaged party demands a compensation for her disadvantage in opportunities, and therefore, is discontent with any payoff which she perceives does not include such compensation. As a result, she is more likely to continue a vendetta. Note that, as mentioned earlier, the loser's distributional concerns must clash with her concerns for equal opportunities in the vendetta game, and may induce *her* to start a vendetta which the winner continues for her disadvantage in opportunities. For the 32.5% of losers who do not start a vendetta, opportunities may weigh more heavily than the distribution of material payoffs (which, one should duly notice, are not realized at that moment yet).

payoffs and are hence *diverse* (Pattanaik and Xu, 2000). Contrary to Sugden and Wang's claim (p. 5 and 6), our concept *does not* rule out (strictly) dominated strategies.[4] To see this immediately, take the two procedures from (Chlaß et al., 2019) in Appendix C.1. For the responder in our mini ultimatum game, we count four effective opportunities, see Appendix C.2. Clearly, the count includes the strictly dominated strategy to always reject. This is true for all players and also holds for weak dominance. A substantial difference with our concept is, however, that we begin with players' *preferences*. Sugden and Wang use *material payoffs* instead and assume that, when these are distinct, the decision maker will think the options entailing them expand her freedom to choose. This need not be. Take a decision maker with reciprocal preferences. She may not derive additional utility from yet another kind option when she already has a plethora of ways to be kind; an additional unkind option when she has none so far, may, on the other hand, be eligible to her set of effective opportunities. Note that such distinctions are impossible with Sugden and Wang's measurement of opportunities. Note, too, that since we begin with players' preferences, we can construct sets of opportunities for a predefined set of economic preference models. The concept also lends itself nicely to assessing opportunities in terms of *merit*: An opportunity which is earned is preferred over other options by its value *minus* the disutility from earning it (Chlaß and Moffatt, 2012/2017, p. 24). It causes a lesser, or no advantage in freedom of choice than an unearned one, and is then more acceptable. Contrary to the authors' claim (p. 5) therefore, our concept leaves decisive room for a reference to *the nature* of options. My comment closes with Chlaß (2010) which studies whether players compensate inequality of opportunity behaviourally. In a game which distributes decision and information rights asymmetrically, I analyze parties' behavioural reactions to the asymmetry by regressing equilibrium deviations on two instruments for the intrinsic valuation of equal opportunities (Chlaß et al., 2019, Chlaß and Riener, 2015/2025). If research questions, hypotheses, and terminology are similar to Sugden and Wang's, note that the study was available as a conference paper before theirs was conducted.[5]

2 A reversed Akerlofian lemons problem

The game in Figure 1 is a common value auction which reverses the information asymmetry in Akerlof's lemons problem. It assigns the information advantage to a buyer who makes an offer to a seller that does not know the quality of her own commodity. I wrote the game originally for a study on the winner's curse, i.e. (Chlaß, 2011), such that the potentially cursed party would move last and may condition her action on an event which has already unfolded, rather than on a hypothetical event taking place in the future (Charness and Levin, 2009). Along the way, I noticed that the new game assigned decision and information rights which we had formalized for a separate study at the time, asymmetrically across players: the first mover has greater decision and information rights than the second. The buyer chooses from a continuous set of offers whereby the seller has two choices only: to accept, or to reject the offer. The buyer knows the quality of the commodity but does not know how the seller will choose in the future; the buyer therefore

[4] The strictly dominated option would, for a narrow self-interested decision maker, simply enter the set of effective opportunities *first*.

[5] and presented in an invited seminar I gave at the University of East Anglia on Dec 16th, 2010, see `https://web.archive.org/web/20110529065530/http://www.uea.ac.uk/eco/news`, where I discussed it with Bob Sugden.

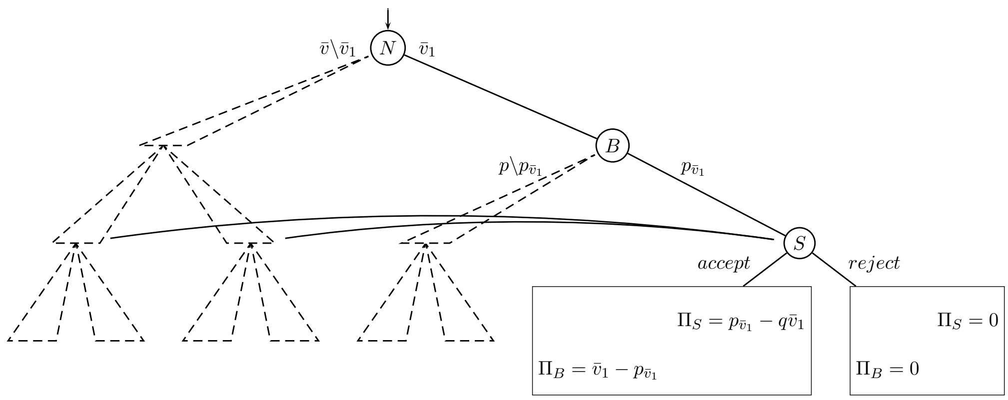


FIGURE 1: THE INFINITE GAME TREE FOR AKERLOF'S LEMONS PROBLEM WITH REVERSED INFORMATION ASYMMETRY; ONE EXEMPLARY PATH OF PLAY.

distinguishes the terminal histories of the game by quality, but for each quality, two terminal histories are bound together. The seller does *not* know the quality (Nature's move) and does therefore not distinguish any of the terminal histories. One might argue that Sugden and Wang only seek to address players' opportunities but not their information rights. Note, however, that the two are typically interdependent.[6] Our subsequent work (Chlaß and Riener, 2015/2025), see Appendix D, disentangles decision and information rights and finds no evidence that players seek to remove or compensate advantageous asymmetries in information rights. Here, I nevertheless aim to privilege one party in a consistent way by both types of rights.

In order to illustrate the game, I briefly go along the one exemplary course of play highlighted in Figure 1: Nature N moves first and draws the value of the seller's commodity, i.e. $\bar{v}_1 \in [0, 10]$. Next, buyer B makes offer $p_{\bar{v}_1} \in [0, 10]$. Note that her information set at node B contains only one element, and that she is therefore informed about Nature's draw $\bar{v}_1$. Finally, seller S decides to accept or reject the offer. Her information set is nonsingleton, and she is therefore not informed about Nature's draw. For every way in which the game may unfold, B has a continuous set of options to choose from; S only has two. If S accepts B's offer, she has payoff $\Pi_S = p_{\bar{v}_1} - q \cdot \bar{v}_1$, if she rejects, $\Pi_S = 0$. Thereby, q with $q \in [0, 1]$ denotes the well-known common value parameter which links buyers' and sellers' valuations of the commodity; note that the seller valuating the commodity less is a necessary condition for trade, and that the smaller the difference in valuations, the larger is Akerlof's social dilemma. Altogether, $\Pi_S = I \cdot (p_{\bar{v}_1} - q \cdot \bar{v}_1)$ where indicator function I takes on a value of One if S accepts. B has payoff $\Pi_B = I \cdot (\bar{v}_1 - p_{\bar{v}_1})$.

A Perfect Bayesian Nash Equilibrium can be found. All offers made under common knowledge of rationality imply $\bar{v} \geq p$. A risk-neutral seller updates her belief about the

[6] In a simultaneous game, the player with fewer opportunities lacks information about more contingencies than her opponent (she knows her own choice but not her opponent's). To see this, it is helpful to depict the simultaneous game in extensive form, let the player under consideration move first, and count the terminal histories she distinguishes when called upon to play. Sequentiality introduces the complication that the party who moves first does, by construction, not know her opponent's future actions. With *perfect information*, her opponent *knows* the history of the game, distinguishes more terminal nodes and has the advantage in information. By introducing *imperfect* information, the game in Fig. 1 repairs this drawback: the player who moves first has greater decision rights; she may not know her opponent's moves, but knows nature's move which her opponent does not. As a result, the party who moves first has greater decision *and* greater information rights: she distinguishes more terminal histories than her opponent.

quality of the commodity, obtains $E(v|v \geq p)$, and computes her own expected payoff which is only nonnegative if the offer is at least $q \cdot E(v|v \geq p)$. The buyer cannot improve by making a lower bid to signal lower qualities. To see this, suppose she were to make some offer d with $0 \leq d < E(v|v \geq p)$. The seller would then expect to make a loss on expectation, reject, and the buyer would have been better off to offer more. Signalling a higher quality would not leave the buyer better off either. Suppose she were to make an offer larger than $q \cdot E(v|v \geq p)$. In this case, the seller would also accept, but the buyer would have improved by offering less, i.e. by lowering her bid until $q \cdot E(v|v \geq p)$.

If a player holds a preference for equal opportunies, she is willing to forego some payoff to bring this equality about. If she must leave the inequality in place and cannot rewrite the rules of the game, she concedes payoff to the opponent, compensating the latter for her lesser freedom to choose. It stands to reason that the player can only afford to do so if she also has the respective financial means. A buyer who seeks to compensate a seller aims at offering a higher price; to what extent she can raise her offer will depend on the value of the commodity $\bar{v}_1$. Sellers in turn who hold a preference for equal opportunities will only accept offers which they deem include such a compensation. The minimal offer they are willing to accept increases. Again, this increase will depend on how much of a compensation a seller expects the buyer can actually afford. I study three situations; one with frequent low qualities; one with intermediate qualities; and one with frequent high qualities. Since one needs intuitive statistical distributions which draw values from the same interval, I assumed a triangular distribution with mode 0, i.e. $S[0, 10, 0]$, a uniform distribution $U[0, 10]$, and a triangular distribution with mode 10, i.e. $S[0, 10, 10]$. Note that $S[0, 10, 0]$ and $S[0, 10, 10]$ concentrate more probability around their expected values than $U[0, 10]$.

Corollary 1. *If Nature draws* $f(v) = U[0, 10]$, $p_{PBNE} = \frac{10q}{2-q}$ *for* $\bar{v} \geq \frac{10q}{2-q}$ *and* $p_{PBNE} = [0, \bar{v}] \quad else.$

Corollary 2. *If Nature draws* $f(v) = S[0, 10, 0]$, $p_{PBNE} = \frac{-10q}{-3+2q}$ *for* $\bar{v} \geq \frac{-10q}{-3+2q}$ *and* $p_{PBNE} = [0, \bar{v}] \quad else.$

Corollary 3. *If Nature draws* $f(v) = S[0, 10, 10]$, $p_{PBNE} = \frac{1}{2}\frac{30-20q-10\sqrt{9+12q-12q^2}}{-3+2q}$ *for* $\bar{v} \geq \frac{1}{2}\frac{30-20q-10\sqrt{9+12q-12q^2}}{-3+2q}$ *and* $p_{PBNE} = [0, \bar{v}] \quad else.$

Proof. All three proofs are found in Appendix B. □

Remark. A buyer who holds a purely procedural preference and seeks to compensate the seller for their lesser rights, raises her offer by λ_S; a seller who demands such a compensation for the asymmetric way in which the game treats her, raises her acceptance threshold by λ_B whereby $\lambda_{A,B}$ depend on the degree to which players care for the equality of rights. The buyer's financial means to offer compensation increase, on average, from $S[0, 10, 0]$ to $U[0, 10]$, to $S[0, 10, 10]$. Note, however, that the amount of resources is known with different precision: In $S[0, 10, 0]$ and $S[0, 10, 10]$ sellers' beliefs about their own quality, and therefore, about the resources available to the buyer, have greater precision than for $U[0, 10]$ where qualities vary greatly around the expected value. In short, for $S[0, 10, 0]$, sellers are relatively sure the buyer has few resources, for $S[0, 10, 10]$, they are relatively sure the buyer has ample resources to compensate. The solution for general triangular distributions is provided in (Chlaß, 2010).

So far, I have looked at fully rational players with purely procedural preferences; note, however, that a seller may not update her beliefs at all, or only partially, under information asymmetry (Eyster and Rabin, 2005). In this case, she may accept from $q \cdot E(v)$ onwards similar to a fully cursed player with $q \cdot E(v) = q \cdot \frac{10}{3}$ for $S[0, 10, 0]$, $q \cdot E(v) = q \cdot 5$ for $U[0, 10]$, $q \cdot E(v) = q \cdot \frac{20}{3}$ for $S[0, 10, 10]$, and add the compensation she requires to this amount. The buyer may make an offer under the assumption that the seller does not at all, or only partially, update their beliefs if she is just as ignorant of the information her offer emits, as the buyer is ignorant of receiving it.

3 Instrumentation of purely procedural preferences

In the game at hand, the full range of outcome -, expectation -, and intention - based social and procedural preferences may be at work. We review some forty models of these and the ethical criteria they build on in (Chlaß et al., 2019, pp. 109-115 and online appendix E, pp. 13-17; Chlaß and Miettinen, 2026, p. 20). Note, too, that if players compensate the asymmetric rules of a game in the game itself, causal inference via the Neyman-Rubin causal model becomes difficult. Take Sugden and Wang (2020)'s setup and assume that the authors had used *two* vendetta games instead, one with greater asymmetry in players' opportunities as treatment intervention, followed by one with less (or preferably, no) asymmetry to elicit players' emotions as response. If $\mathscr{L}$ compensates $\mathscr{W}$ for her lesser decision rights in the first game by taking less and by compensating $\mathscr{W}$ in terms of payoff, then $\mathscr{L}$'s and $\mathscr{W}$'s emotional state after unequal opportunities may either be neutral (adequate compensation), $\mathscr{W}$ may still be upset (inadequate compensation), or $\mathscr{L}$ may be upset because she deems that W requires an unreasonable amount of compensation. Even if the second vendetta game lent itself to measuring emotions without bias – i.e. did not affect emotions itself – the effect of unequal opportunities becomes shrouded when compensation is present.[7]

To solve this issue, I used a causal method for observational data in the context of the game in Figure 1, i.e. an instrumental variable approach developed in our original study. Purely procedural preferences where they express a player's utility from the equality of rights, were linked to a then new ethical criterion: social contract reasoning as coined by Lawrence Kohlberg in his *Kohlberg class 5*, see e.g. (Kohlberg and Hersh, 1977/2009); I will focus on this criterion in detail later on. Subjects' use of the criterion in their moral judgement was obtained by means of Georg Lind's (1977/2021, 2008) moral judgement/competence test (M-J/C-T) which has important conceptual advantages from an Economist's point of view over its direct and only competitor, i.e. James Rest's (1977) Defining-Issues-Test (D-I-T).[8] Note, too, that because the M-J/C-T began and continued as a German enterprise before being validated in other cultures, I put greater trust in its cultural validity for a sample of Germans.

To briefly review the instrument, the first step is to revisit its correlation with individuals' purely procedural preferences. In (Chlaß et al., 2019), we designed a yes-no and an ultimatum game, each with the same payoffs such that *given* identical pure strategies

[7] To illustrate, a 'causal effect' (shrouded by compensation and learning) would be measured as the difference in emotions between this and a second treatment with two subsequent symmetric vendetta games.

[8] The D-I-T can easily be completed from various perspectives, i.e. is manipulable in a straightforward way (Bebeau, 2002) and would therefore be sensitive to attempts at ex-post rationalization or at answering in a socially desirable way. In contrast, I provide evidence on the *exogeneity* of Kohlberg class 5 elicited via the M-J/C-T in section 7.

and identical pure strategy beliefs in both games, players were predicted to be indifferent *in*, and *off equilibrium* by the forty outcome -, expectation -, and intention based models mentioned. The two games are shown in Appendix C.1. In each game, a proposer may offer a fair or a generous proposal; a responder may accept or reject, in which case both parties earn zero. The yes-no game proceeds simultaneously, the ultimatum game sequentially with perfect information. The yes-no game introduced equal rights; the ultimatum game unequal ones, favoring the responder. We observed relatively few subjects who chose the same pure strategies and held identical pure strategy beliefs in the two games *and* were willing to pay some amount for either game. Note that since our inequity aversion over rights introduced two parameters, one by which a subject dislikes being worse off (having lesser rights), and another by which she dislikes being better off (having greater rights), she may prefer either game. The small set of revealed preferences linked to social contract reasoning as did a slightly larger set of stated preferences, see Appendix C.3. Interestingly, however, the 47% of subjects who reasoned most strongly in terms of the social contract *always* chose procedurally variant strategies and/or held procedurally variant beliefs. We therefore suspected that subjects' decisions within each procedure already accounted to some extent for the distribution of rights, and that subjects expected their opponents' decisions to do so, too. In order to see whether this was indeed the case, we matched subjects on their beliefs and actions, and, for all clustered sets of strategies and beliefs, regressed subjects' choices of the game on their *Kohlberg 5* scores. Throughout, subjects' behaviour against their self-interest could indeed be statistically explained by social contract reasoning, see Appendix C.5. In that first study, one and the same game accorded players equal decision, *and* equal information rights, and we could not infer which of the two mattered more. In (Chlaß and Riener, 2015/2025), we therefore studied pairs of games in which an individual could determine *either* her opponent's decision rights *or* her opponent's information rights while payoffs were identical, see Appendix D.1. Only when individuals had to choose how many decision rights to accord their opponent, did we observe willingness to forego payoff. Subjects' choice of equal decision rights, and their altruism under unequal decision rights, could both be explained by social contract reasoning, i.e. *Kohlberg class 5*, see Appendix D.4. We also changed payoffs such that decision rights would not be at stake, and quite rightly, observed hardly any willingness to forego payoff.

The next step was to understand the *conditions* under which social contract reasoning can instrument purely procedural preferences. The appendix to our original paper provided data from four subsequent studies which look at other variables that might have caused the link described above; three of these studies use the same subject pool and the same (standardized German) version of the M-J/C-T. We noted that social contract reasoning correlated with subjects' *field of study: Law*, and their *gender*. Results of a (new) fourth study (Chlaß and Miettinen, 2026) are found in Online appendix F.[9]

In the original study, we observed a second, *negative* correlation of purely procedural preferences with a facet of moral judgement by which the use of social contract reasoning depends on one's peers opinions, social norms, or one's social image, i.e. an interaction of *Kohlberg classes 3 & 4* and *Kohlberg classes 5 & 6*. Since for this variable, we also provided results on potential latent correlates, I use it as a redundant second instrument

[9]The study analyses in particular the relation between moral judgement and personal values. Value *achievement* correlates negatively with nearly all moral preferences. The instrumentation remains valid, however, since *achievement* correlates negatively with both *Kohlberg 5* and the second instrument whereas only *Kohlberg 5* correlates positively, the other instrument negatively, with purely procedural preferences.

here, to afford the first instrumentation more credibility. Scores for social contract reasoning and all other *Kohlberg classes* are computed the same way as in (Chlaß et al., 2019, Chlaß and Riener, 2015/2025, Chlaß et al., 2023, Chlaß and Miettinen, 2026).

4 Meaning and measurement of the instrument

The moral judgement/competence test[10] by which I elicit subjects' use of social contract reasoning is based on work in developmental psychology with a longstanding tradition pioneered by Jean Piaget (1948). Following up on this work, Lawrence Kohlberg conducted extensive field work to identify which ethical criteria individuals use to make moral judgements, i.e. to derive whether a course of action is ethically right or wrong, and how the use of these criteria develops throughout the individual lifespan. The ethical criteria, observed mainly in large-scale interviews, were divided into a taxonomy of six categories. The M-J/C-T confronts subjects with ethical criteria and moral problems taken from this field work; each item of the test loads on one category of moral judgement. Note that Kohlberg advocates an ordinal logic between these categories, that is, an improvement in the quality of moral judgement from category one to six. Our analysis did not assume such a logic and also does not typify subjects by the type of criterion they prefer most; it measures individuals' use of all six classes of ethical criteria. In *Kohlberg class 1 & 2*, individuals infer from material punishment and reward whether an action is right or wrong. *Kohlberg class 3* lists criteria which, in some way, infer from interindividual relations with one's immediate group whether an action is ethically right or wrong. These criteria are, for instance, the expectations of one's peers, one's social image with the latter, or social norms. In *Kohlberg class 4*, the individual acknowledges the relevance of norms from outside her immediate group, such as the law, the authorities enforcing it, and the relevance of these norms for the workings of society. In *Kohlberg class 5*, individuals acknowledge and refer to the mutuality of a democratic social contract and the equality of rights stipulated therein. *Kohlberg class 6* lists universal individual principles of conscience such as Kant's categorical imperative, the concept of inalienable human rights; freedom of choice, the individual will and dignity to which individuals refer when judging whether a course of action is ethically right or wrong. In our previous paper, we located economics' social and procedural preference models within Kohlberg's six classes; for a completed overview, see Appendix H. The idea was that, by means of the M-J/C-T, it is in principle possible to elicit the respective normative element which each preference model assumes the individual to intrinsically care for, and which the model therefore assumes to cause the individual's departure from narrow rational self-interest. By understanding which ethical criteria explain subjects' decisions, one can infer which types of preference may be at work.

The M-J/C-T introduces two moral problems, or dilemmas, in which all courses of actions portrayed may be ethically right or wrong by some ethical criterion – hence the word *'dilemma'*. The first problem portrays workers breaking into a factory stealing evidence that management was listening in on them; the second portrays a doctor that assists a woman who is terminally ill, to commit suicide upon her request. Subjects first submit their opinion on whether the action was right or wrong. Subsequently, subjects rate by how much they would agree or disagree to use a given ethical criterion to judge the action was <u>right</u> – one criterion for each of the six *Kohlberg classes*. Next, the test lists

[10]The test is freely available for research purposes at http://www.moralcompetence.net. An excerpt is given in online appendix D with kind permission by its author, Georg Lind.

another six criteria, and asks subjects by how much they would agree to use each in order to derive the action was wrong. Subjects submit their answers on a 9 point Likert scale from -4 (do not agree) to +4 (fully agree). For the second dilemma, the test proceeds analogously. Altogether, the test therefore introduces four items for each Kohlberg class. A subject's preference for each class is the average over her four responses. Throughout all studies mentioned, this average is adjusted for subjects' personal use of the Likert scale, i.e. for the range between the most extreme values of the Likert scale a subject ever ticks in the entire test. For details, see e.g. (Chlaß et al., 2019, Chlaß et al., 2023).

5 Experiment

286 students participated, recruited via ORSEE (Greiner, 2015), randomly drawn from all fields of study at the University of Jena. Subjects who had participated in (Chlaß, 2011) which first studied the game in Figure 1, were not invited. The experiment was conducted at the laboratory of the Max Planck Institute of Economics in Jena and was programmed in zTree (Fischbacher, 2007).

At the appointed time, the laboratory opened and participants drew their cabin number randomly from a box with face down, well shuffled cards. Each participant seated herself in the cubicle corresponding to the number she had drawn; the laboratory seated 32 participants. In each cubicle, a cover page with general information was placed face-up on each keyboard, explaining that the experiment would proceed in three independent parts. Cover page, instructions, control questions, and supplements to the instructions are found in online appendices A and B.

Five minutes into the session, the zTree programme was started, guiding participants to and through part One. In part One, subjects submitted their risk-preferences by choosing between lotteries in a Holt-Laury style format in order to generate some income since common value auctions typically generate losses, and to avoid that loss aversion or solvency concerns would shroud subjects' play.[11] Note that risk preferences neither correlate with the instrument(s)[12], i.e. *Kohlberg class 5* or con×post, nor with buyers' response variable, i.e. offers. For the seller side, where risk preferences can affect the response variable, with risk aversion slightly increasing sellers' willingness to accept an offer, I conduct a separate control treatment. Eliciting risk preferences before, or after the game, does not affect their acceptance behaviour. There is therefore no evidence that this procedure in three parts has brought about the results I report later on, or affected them in any way. Individuals chose between lotteries at their own speed. Part One completed, no feedback was given, and the zTree programme guided participants to turn the instructions for part Two which had been placed face down out of sight at the top right of the cabin desk. Having completed reading the instructions, subjects pressed an OK button, and a series of control questions appeared. Afterwards, participants were randomly assigned buyers or sellers for part Two and played 20 rounds of the game in Fig. 1 in a random strangers matching. After each round, participants received feedback on the

[11] The show-up fee as per the general rules of the subject pool at the time was €2.50, too little to account for losses. The research group kept the fee low such as to avoid experimenter demand effects, i.e. to avoid 'cooperation' on the part of the subjects with the experimenter, or niceness toward other subjects.

[12] The online appendix to (Chlaß et al., 2019), available from https://www.chlass/research provides the respective null results. Risk aversion has, in particular, a null effect on the instruments for the data from (Chlaß and Riener, 2015/2025) (see also online appendices G.3.2 and G.3.3 to this paper), the current study, and, as shown in Online appendix F to this paper, also for (Chlaß and Miettinen, 2026).

buyer's offer, whether or not the seller had accepted that offer, and their round profits. The feedback did not inform participants on their overall profits. A session introduced either $S[0, 10, 0]$, $U[0, 10]$, or $S[0, 10, 10]$. Otherwise, each session proceeded the same way. For ten rounds, common value parameter was set to 0.2, for another ten rounds, to 0.6. Each session with 32 subjects[13] was divided into two matching groups of 16 subjects; one starting out with ten rounds of common value parameter 0.2, the other with 0.6. Throughout, subjects could use a simple calculator shown on each screen. For part Three, the zTree programme was halted, and subjects completed the M-J/C-T on paper, after which the programme resumed with Hans-Jörg Eysenck's P-E-N-L personality inventory – the latent variable to which I relate the winner's curse in the companion paper for which the game in Fig. 1 was originally written. I elicit the inventory here to avoid omitted variable bias. Finally, subjects received information on their overall payoff from all three parts. Buyers' average payoffs were €17.54 ($25.78 at the time) with €15.31 for $S[0, 10, 0]$, €17.56 for $U[0, 10]$, and €19.74 for $S[0, 10, 10]$. Sellers' average payoffs were €13.10 ($19.26 at the time) with €12.50 for $S[0, 10, 0]$, €13.79 for $U[0, 10]$, and €14.65 for $S[0, 10, 10]$. A session lasted approximately 90 minutes. For $U[0, 10]$, there were 94 participants; for $S(0, 10, 0)$ 96 participants, and for $S[0, 10, 10]$ also 96 participants. For each probability distribution, there were six independent matching groups.[14]

6 Results / 6.1 Descriptives

Figure 2 mirrors the distribution of bids (turquoise) against the distribution of accepted bids (green) for all treatments, grouped by the payoffs which these accepted bids entail. It also plots a density for each distribution. 'Mutual gain' denotes nonnegative payoffs for both parties; 'loss' a negative payoff for the seller but not the buyer, and 'sacrifice' a negative payoff for the buyer but not the seller. The two red dashed lines demark a seller who does not update her beliefs at all (left-hand line) and a seller who fully updates and accepts only $p_{PBNE,v\geq p}$ (right-hand line). In between, we find all bids which sellers who only partially update, would accept. One sees upon first glance that, whatever the treatment, sellers incur losses rather seldomly – in maximally 8.4% (40 out of 480) interactions. This might simply confirm Charness and Levin's (2009) idea that the timing of events is critical to people's strategic thinking; that individuals do see the point of accounting for an event which has already unfolded, but not for a future event. In the companion paper (Chlaß, 2011), however, where an algorithm mimics a rational buyer who believes in the seller's rationality, and expects the seller to believe in the buyer's rationality, I observe seller losses in 17.21% (413) of all 2400 interactions for $S[0, 10, 0]$,

[13] One session with 30 participants only had matching groups with 16 and 14 participants.

[14] In the descriptives section, I will refer to the robot treatments of the companion paper. Therein, human participants play as sellers against a computerized buyer which offers the $\text{PBNE}_{v\geq p}$ price when mutually beneficial trade is, in expectation, possible and randomizes otherwise amongst the rational offers it can make. In these treatments, there are 96 sellers for $S[0, 10, 0]$, $q = 0.6$ who play 25 rounds; 95 sellers for $U[0, 10]$, $q = 0.6$, 63 of which play 20 rounds and the remaining 32 of which play 25 rounds. These robot treatments first elicit personality, then risk preferences, and then introduce Akerlof's reversed lemons problem from Figure 1. Another treatment for $U[0, 10]$, $q = 0.6$ with 78 participants first introduces the game, mentioning the fixed payoffs for the personality inventory as well as minimal and maximal payoffs from the risk preferences (without reference to what these questionnaires are about) to keep the perceived endowment constant, and then elicits risk preferences and personality. The winner's curse does not depend on whether risk preferences and personality are elicited after the game instead of before, see footnote 17.

and in 22% (451) of all 2060 interactions for $U[0, 10]$, see Appendix E.1. Appendix E.2 provides key figures for robot and human treatments.[15]

Either therefore, buyers in Figure 2 make higher offers than the algorithm in the companion paper; or sellers have higher acceptance thresholds; or both. It is in particular possible that buyers offer compensations for the asymmetry of rights to sellers, and that sellers have higher acceptance thresholds since they require to be compensated. The mechanism at work in Figure 2 is of particular interest: already strong at the outset, it gains further strength as the experiment progresses. In the final five rounds, losses drop to 8.4% from 14.2% in the first five rounds for $S[0, 10, 0]$ and $q = 0.6$; to 2.6% from 16.7% for $U[0, 10]$ and $q = 0.6$, and to 8.3% from 10.8% for $S[0, 10, 10]$ and $q = 0.6$. In the companion paper, such dynamics do not unfold. Losses persist with 18.3% in the first five, and 15.2% in rounds[16] 16 to 20 where $S[0, 10, 0]$ and $q = 0.6$. For $U[0, 10]$ and $q = 0.6$, I observe 23.4% seller losses in the first five rounds, and 21.1% in rounds 16 to 20. Note that the companion paper switches off (nearly) all social preferences, and in particular preferences for equal opportunities. Whatever empirical mechanism is at work once human buyers and sellers interact, overlays the winner's curse phenomenon, and increasingly so with experience.[17] Next, I regress buyers' offers on the instruments *Kohlberg class 5* and con×post, controlling for critical latent variables.

6 Results / 6.2 Instrumentation Buyers

A buyer deviates from rational self-interest, if she <u>*could*</u> rationally – without incurring a loss – make the smallest price offer which a rational self-interested seller would accept, but chooses not to offer this price. This is true for $q \cdot E(v|v \geq p) < \bar{v}$ and $p \neq q \cdot E(v|v \geq p)$: either the offer falls short of, or exceeds, the equilibrium offer. If, in this case, the buyer adds a premium to her offer to compensate the seller's inferior position of rights, she offers $p > q \cdot E(v|v \geq p)$ given $q \cdot E(v|v \geq p) < \bar{v}$. If the buyer ignores the informa-

[15] Since the companion paper sought to explain the winner's curse by personality traits, it only introduced distributions which make small qualities v and therefore, the winner's curse phenomenon, likely: the original uniform distribution $U[0, 10]$, and the triangular distribution with mode zero $S[0, 10, 0]$. Common value parameter q was $q = 0.6$ throughout.

[16] The companion paper has twenty-five rounds, this paper twenty. In order to compare a similar amount of experience, I use data from rounds 16 to 20.

[17] For the robot treatments, I test in particular whether losses in the game of Figure 1 are more frequent when eliciting risk preferences before, or after the game. Looking at independent first round data, the frequency of losses does not differ significantly across these two conditions by Fisher's exact test, *p-value* = 1. Moreover, each robot treatment (eliciting the game first, or last), differs the same way from the human game: first-round losses are more prevalent with robot buyers, i.e. by *Fisher's exact test, p-value* = 0.034 (robot/personality and risk <u>first</u> vs. human), and *p-value* = 0.029 (robot/personality and risk <u>last</u> vs. human). Seller decisions which imply losses *in expectation*, are also similarly frequent in both robot conditions, i.e. *p-value* = 0.289. In each robot condition, such decisions are more likely than with two human players, i.e. by *Fisher's exact test, p-value* = 0.015 (personality and risk first vs. humans), and *p-value* = 0.003 (personality and risk last vs. human). If I supplement the human treatment data with those matching groups who encounter their first round with $q = 0.6$ in round 11 of the experiment, I obtain the same test results. It is noteworthy that *risk preferences* differ significantly if elicited upfront as compared to after personality, i.e. Wilcoxon's Exact Rank Sum test, *p-value* = 0.039 and *p-value* = 0.004. Seller behaviour in contrast (minimally acceptable offers, independent first round data) does not change if risk preferences and personality have been elicited before the game as compared to afterwards, by Wilcoxon's Exact Rank Sum test, *p-value* = 0.275.

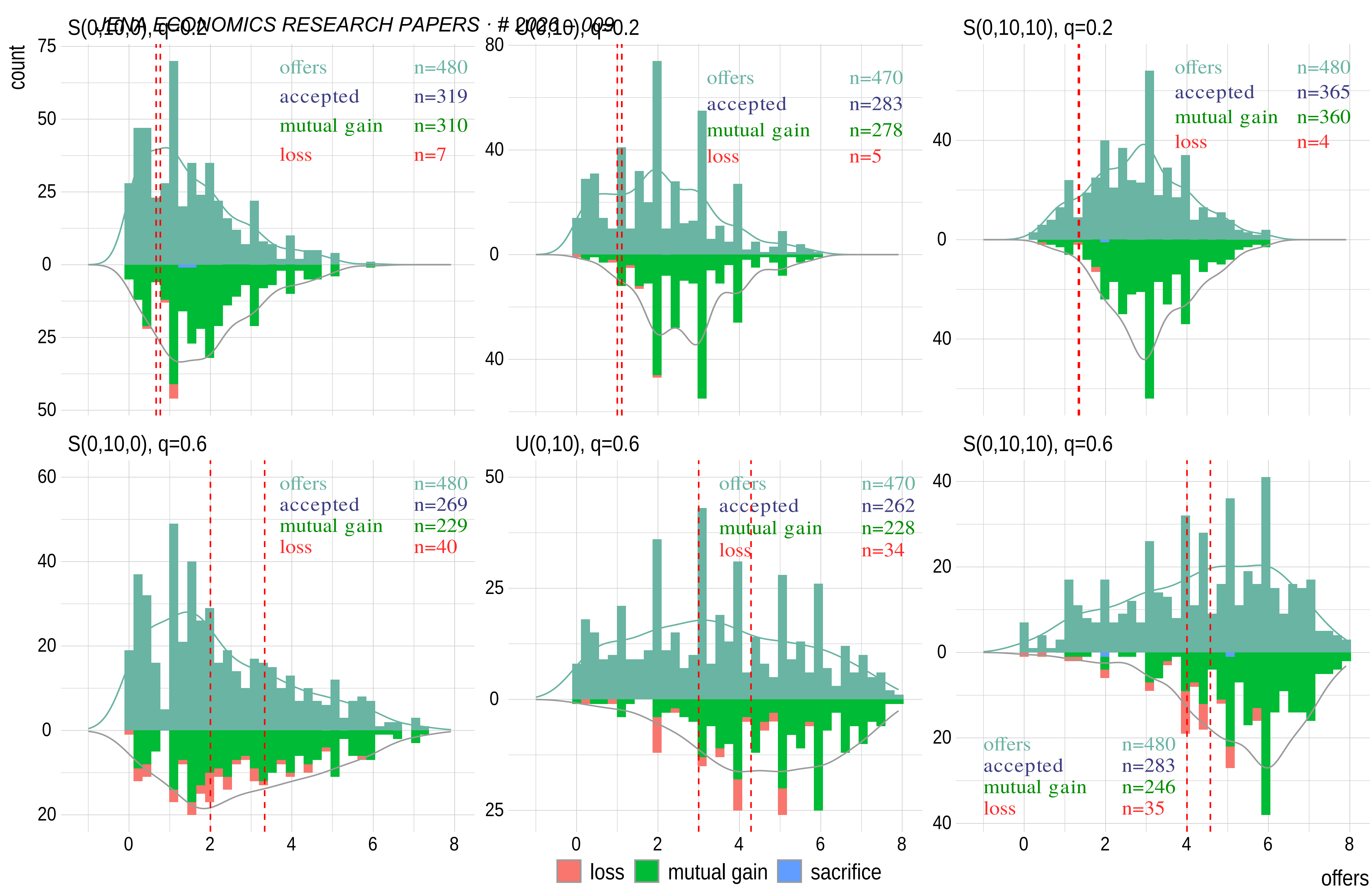


FIGURE 2: OFFERS, MIRRORED AGAINST ACCEPTED OFFERS GROUPED BY THE PAYOFFS THEY ENTAIL. ACCEPTED OFFERS ARE THE SUM OF OFFERS ENTAILING A MUTUAL GAIN, A LOSS, AND A SACRIFICE.

tion emitted by her offer[18], she may add that premium to the smallest offer which *she believes* the seller can accept without making losses on average, that is, to a fully or partially cursed acceptance threshold, in which case $p < q \cdot E(v|v \geq p)$ for $q \cdot E(v|v \geq p) < \bar{v}$. All empirical deviations from rational self-interest where a buyer's bid falls either short of the equilibrium offer, or exceeds the equilibrium offer when the latter could rationally have been made, are clustered via Eucledian distance by Ward's minimum variance technique (Ward, 1963) – a standard – to obtain three homogenous groups of similar size. These are ranked by their group mean which yields three ordinal categories.[19] The first category has group mean −0.68 with 488 deviations; the second category group mean 0.55 with 630 deviations, and the third category has group mean 2.13 with 946 deviations. I model these three ordinal categories by means of an ordinal logit model with standard errors clustered at the individual, and at the matching group level. Cut-off points between the three categories are highly significant. Table 1 shows the fitted model. Specification 1 includes all variables; specification 2 only significant variables from specification 1 such as to reduce efficiency losses; the third specification shows that critical effects do not depend on controls. Specification 2, the efficient model without apparent omitted variable bias, serves as reference for the text. Note, however, that all three specifications show similar results.

The instrument *Kohlberg class 5* has a throughout significant positive coefficient which is 0.310 with *p-value* = 0.028 in the full, 0.323 with *p-value* = 0.000 in the efficient, and 0.328 with *p-value* = 0.001 in the small model. Looking at marginal effects which are not displayed in Table 1, the instrument shifts probability to the highest category; *Kohlberg class 5* has a marginal effect of −0.041 with *p-value* = 0.000 on category one, −0.015 with *p-value* = 0.001 on category two, and a marginal effect of 0.057 with *p-value* = 0.000 on category three. *Kohlberg class 5* therefore increases the likelihood of those bids which lie farthest above the equilibrium offer. The second instrument *con×post* also has the correct (negative) sign, i.e. −0.261, *p-value* = 0.020, and also implies that stronger purely procedural concerns shift likelihood toward category 3, i.e. by 0.046, *p-value* = 0.021.[20] Note that this effect still holds when the missing component of interaction *con×post* is added to (2); after adding *con* which averages *Kohlberg class 3* and *4*, the effect on category 3 is 0.061, *p-value* = 0.057, and *Kohlberg class 5* still has marginal effect 0.057, *p-value* = 0.002. To conclude, both instruments indicate that buyers who hold purely procedural preferences, add a premium to their offers.

6 Results / 6.3 Instrumentation
Sellers

A seller deviates from rational self-interest if she *accepts* offers which, in expectation, yield her a loss, i.e. $p < q \cdot E(v|v \geq p)$, or if she *rejects* offers which, in expectation, yield her a gain, i.e. $p > q \cdot E(v|v \geq p)$. If a seller seeks compensation for her inferior

[18] the information which a seller who believes in the buyer's rationality, and in the buyer's belief in her own rationality, would *infer from* the offer, whether or not her rationality beliefs are accurate.

[19] In principle, deviations from equilibrium are cardinal and differences between categories can be measured. Note, however, that in some regions, there are only few observations whereas in others, there are many. Grouping the cardinal variable into similarly sized categories arranges the matter and accounts for the fact that small deviations from equilibrium may simply be mistakes.

[20] To save technical effort, I do not correct the estimations for the imperfect correlation of the instrument with the instrumented variable. Reported *p-values* are hence (very) conservative.

Dependent Variable: Buyers' small, intermediate, and large Deviations from rational Self-Interest			
x_j ↓ specification →	(1)	(2)	(3)
Kohlberg class 1	−0.092 (0.128)		
Kohlberg class 2	-0.332^b (0.147)	-0.284^b (0.114)	-0.238^b (0.100)
Kohlberg class 3	0.056 (0.188)		
Kohlberg class 4	0.158 (0.152)		
Kohlberg class 5	0.310^b (0.141)	0.323^a (0.088)	0.328^a (0.102)
Kohlberg class 6	0.248^b (0.107)	0.243^b (0.089)	0.224^b (0.091)
con×post	−0.335 (0.213)	-0.261^b (0.113)	-0.256^c (0.144)
quality $\bar{v}$	0.593^a (0.043)	0.575^a (0.038)	0.544^a (0.035)
Period	0.039 (0.042)		
Risk aversion	0.061 (0.058)		
Age	−0.012 (0.029)		
Gender:female	0.390^c (0.217)	0.384^b (0.188)	
Extraversion	−0.028 (0.024)		
Neuroticism	-0.053^a (0.019)	-0.041^b (0.017)	
Psychoticism	0.044^c (0.025)	0.040^c (0.024)	
Business/Economics	0.413^c	0.437^c	
Law	-0.739^c	-0.692^c	
Social and Behavioral Sciences	0.050		
Education	−0.023		
Engeneering	−0.655		
Philosophy	0.260		
Theology	-1.539^a	-1.675^a	
Count R^2	0.603	0.580	0.568
observations	2064	2064	2064

Table 1: Buyer deviations from rational self-interest, explained by instruments *Kohlberg 5* (positively related with a preference for the equality of rights) and *con×post* (negatively related with that preference).

Notes. Significance levels are indicated by $a : p < .01$, $b : p < .05$, $c : p < .10$. Table 1 displays coefficients of ordinal Logits of buyer departures from rational self-interest with standard errors clustered twoways at the individual and at the matching group level. Cut-off points (equivalent to the intercepts of the upper categories) for specification (2) are: category 1→2: 2.420 (std. error 0.498) and 2→3: 4.344 (std. error 0.471).

position of rights, she increases the minimal offer she is willing to accept by the compensation she requires. The latter depends on the strength of her purely procedural preference. If the seller does not update her beliefs about the quality of the commodity by $\bar{v} \geq p$ – the information which every rational offer transmits, see footnote 18 – she may add the compensation to the smallest offer which *she believes* she can accept without making losses on expectation. Recall that sellers do not know the quality of the commodity; they only have information about the distribution of qualities to determine how much compensation buyers can afford.

Buyer offers censor seller deviations in a natural way: a seller can only deviate from rational self-interest if and to the extent by which an offer allows her to do so – that is, if and to the extent by which an offer falls short of, or exceeds, the equilibrium offer. I therefore group seller deviations by a two-stage WARD clustering technique which matches sellers according to how similar an opportunity to deviate from equilibrium they face. Stage 1 clusters negative deviations only. For $p < q \cdot E(v|v \geq p)$, the dendrogram shows six natural groups; I therefore have the algorithm construct six groups.[21] Each group consists of sellers who accept an offer short of equilibrium, and sellers who, faced with an offer similarly short of equilibrium, accept; each group therefore matches sellers who deviate from equilibrium with sellers who do not deviate, given similar opportunity. The higher the cluster resolution, the more accurate the matching. Appendix F summarizes all clusters, including the rejection rate in each. In each cluster, I regress sellers' decision to accept or reject in a simple binary logit model on the full set of *Kohlberg classes*, sellers' personality traits, and dummies for the distributions of qualities which govern how much compensation for the inequality of rights sellers expect buyers can afford. I also use the potential deviation from equilibrium as explanatory variable, in case the matching was not precise enough. Specifications start by interacting instruments and distribution dummies and are then tested down in a classic top-down strategy, removing insignificant variables to gain efficiency, except for necessary controls. Standard errors are clustered at the individual, and at the matching group level.

Table 2 shows the results for *Kohlberg 5*. In the first cluster *1neg* with 158 observations, offers fall, on average, 2.49 ECU ($\sigma = 0.204$) short of equilibrium. *Kohlberg 5* $\cdot$ $S[0, 10, 10]$ has coefficient -8.624, *p-value* = 0.000 and marginal effect -0.986, *p-value* = 0.000. That is, *Kohlberg 5* significantly decreases the likelihood of a successful bid when buyers can easily afford to compensate the seller's lesser rights but don't do so. Where buyers have few resources to compensate, the marginal effect is smaller: *Kohlberg 5* $\cdot S[0, 10, 0]$ has only -0.143, *p-value* = 0.014. The stand-alone coefficient of *Kohlberg 5* for this regression is 2.320 with *p-value* = 0.001; the marginal effect 0.265, *p-value* = 0.000. $S[0, 10, 10]$ and $S[0, 10, 0]$ have coefficients -2.627, *p-value* = 0.020, and -2.508, *p-value* = 0.003; marginal effects are -0.216, *p-value* = 0.000 and -0.227, *p-value* = 0.000, respectively. Note that these distribution dummies may already capture some part of sellers' reaction to unequal rights. In cluster *2neg* with 209 observations, offers fall, on average, 3.44 ($\sigma = 0.44$) short of equilibrium. *Kohlberg 5* $\cdot$ $S[0, 10, 0]$ has coefficient -1.760, *p-value* = 0.010 and marginal effect -0.163, *p-value* = 0.016. Since only the interaction of *Kohlberg 5* with $S[0, 10, 0]$ turns significant, I also consult the second instrument *con×post* which correlated *negatively* with purely procedural preferences in the original study. From Table 3, we see that *con×post* has coefficient 0.610, *p-value* = 0.002, marginal effect 0.053, *p-value* = 0.001, and does not interact

[21] Lower resolutions, say, of three, keep the first two clusters intact and merge the remaining four. Online appendix E.1 shows the respective dendrogram and the resolution of six clusters from the main text; E.2 shows the dendrogram for positive deviations and a cluster resolution of three clusters from the main text.

with $S[0, 10, 0]$ or $S[0, 10, 10]$. *con×post* is also active in cluster *3neg* where offers fall, on average, 1.209 ($\sigma = 0.135$) short of equilibrium and where *con×post* interacts both with $S[0, 10, 10]$ and $S[0, 10, 0]$. These interactions show coefficients and effects analogous to *Kohlberg class 5* in cluster 1.[22] In cluster *4neg* with 188 observations, offers fall short of equilibrium by an average of 1.79 and deviations stem mostly from treatment $U[0, 10]$. There is no evidence that either of the two instruments affects sellers' decision to accept. In clusters *5neg* and *6neg* where offers fall short of equilibrium by an average of 0.221 ($\sigma = 0.117$) and 0.665 ($\sigma = 0.131$), respectively, *con×post* has marginal effects 0.043, *p-value* = 0.021 and 0.051, *p-value* = 0.022, respectively.

Turning to offers which exceed the equilibrium offer, I keep the analysis brief, opting for a resolution of three clusters shown by the dendrogram in Online appendix E.2. As offers increasingly exceed the equilibrium offer, sellers should deem that compensation has been granted, and that it is therefore increasingly right to accept the offer. Signs of the respective coefficients and marginal effects should reverse. As shown in Table A7, Appendix H, this is indeed the case. In cluster *1pos* with 624 observations, offers exceed the equilibrium offer by an average of 0.56 ($\sigma = 0.288$) and *con×post* has negative marginal effect −0.054, *p-value* = 0.034.[23] For clusters *2pos* and *3pos*, rejection rates are extremely low which requires logits for rare events.[24] In cluster *2pos* with 535 observations, offers exceed the equilibrium offer by 1.56 ($\sigma = 0.267$). *Kohlberg 5·* $S[0, 10, 0]$ and *Kohlberg class 5·* $S[0, 10, 10]$ have coefficients 2.445, *p-value* = 0.040, 1.868, *p-value* = 0.048 and marginal effects 0.145, *p-value* = 0.014, and 0.111, *p-value* = 0.013, respectively. Finally, in cluster *3pos* with 414 observations, offers exceed the equilibrium offer by on average 2.876 ($\sigma = 0.664$). *Kohlberg 5·* $S[0, 10, 0]$ has coefficient 3.364, *p-value* = 0.010 and marginal effect 0.132, *p-value* = 0.024. Power logits confirm the sign reversal for *Kohlberg class 5* in clusters *2pos* and *3pos*. Instruments now interact positively with $S[0, 10, 0]$ and $S[0, 10, 10]$.

7 Discussion

Purely procedural preferences for equal effective opportunities (Chlaß et al., 2019,Chlaß and Riener, 2015/2025) were linked to two variables: *Kohlberg class 5* (positively), and *con×post* (negatively). For buyers in the game of Figure 1 who have greater opportunity, each instrumental variable in its own right indicates that such preferences increase acquirer offers. *Kohlberg class 5* has a positive effect; *con×post* a negative effect, and both instruments therefore consistently support the idea that acquirers compensate sellers through the common value for their lesser rights.

On the seller side with lesser opportunity, *Kohlberg class 5* reduces the likelihood of accepting an offer below equilibrium (winner's curse) as a function of the buyer's resources to compensate, whereby $S[0, 10, 0]$ typically provides sellers with little, and

[22] *Con×post·*$S[0, 10, 10]$ and *con×post·*$S[0, 10, 0]$ have coefficients 4.622, *p-value* = 0.001 and 2.650, *p-value* = 0.001; marginal effects are 0.600, *p-value* = 0.000 and 0.344, *p-value* = 0.000, analogous to the negative interactions of *Kohlberg 5* with $S[0, 10, 10]$ and $S[0, 10, 0]$ in cluster *1neg*. *Con×post* has coefficient −3.020, *p-value* = 0.000 and marginal effect −0.392, *p-value* = 0.000, analogous to the positive effect of *Kohlberg 5* in cluster *1neg*.

[23] On an efficient (reduced) model which still features a row of insignificant controls, i.e. all six Kohlberg classes, Law, gender, and deviation, the effect is −0.059 with *p-value* = 0.011.

[24] Power logits assume that regressors have their maximal effect on the probability of success beyond probability 0.5, and, after recoding successes as failures and vice versa, even beyond probability $1 - e^{-1} \sim 0.63$ (Achen, 2002). The *negative* of the estimated coefficients and marginal effects then represent coefficients and marginal effects on offer *acceptance*.

Dependent Variable: Sellers' decision to accept/reject in clusters *1neg*, *2neg*, *3neg*/offers short of equilibrium — Instrument Kohlberg class 5			
x_j ↓ cluster →	*1neg*	*2neg*	*3neg*
cluster mean →	−2.487	−3.441	−1.209
Intercept	9.331	−1.218	-6.587^{c}
	(6.321)	(4.660)	(3.406)
deviation	1.729	1.855^{a}	-4.087^{b}
	(1.675)	(0.599)	(1.589)
Kohlberg class 5 · S(0,10,0)	-1.250^{b}	-1.760^{a}	−0.666
	(0.553)	(0.679)	(0.838)
Kohlberg class 5 · S(0,10,10)	-8.624^{a}		0.374
	(2.065)		(0.951)
Kohlberg class 5	2.320^{a}	0.885^{a}	0.023
	(0.718)	(0.293)	(0.823)
S(0,10,0): yes	-2.508^{a}	0.725	-1.284^{c}
	(0.854)	(0.698)	(0.715)
S(0,10,10): yes	-2.627^{b}		−1.070
	(1.131)		(0.956)
Period	-0.166^{a}	-0.134^{b}	−0.003
	(0.058)	(0.057)	(0.041)
Risk aversion	0.335^{c}	0.274^{b}	0.256^{b}
	(0.174)	(0.133)	(0.116)
Age	-0.475^{b}	−0.064	−0.109
	(0.188)	(0.125)	(0.094)
Gender:female	2.482^{a}	1.656^{b}	1.107^{b}
	(0.552)	(0.784)	(0.531)
Extraversion	−0.013	0.189^{a}	0.153^{a}
	(0.085)	(0.046)	(0.044)
Neuroticism	0.046;	−0.003	−0.021
	(0.047)	(0.048)	(0.037)
Psychoticism	-0.193^{b}	0.112	-0.104^{c}
	(0.092)	(0.109)	(0.061)
Business/Economics	1.998^{c}	-3.062^{a}	0.485
Law	2.193^{c}	−0.585	−0.192
Social and Behavioral Sciences	2.231^{c}	−0.605	−0.518
Education	1.123	−1.364	−0.537
Engeneering	5.164^{a}		0.233
Medicine	1.863	−0.833	-2.100^{c}
Languages	3.126^{b}	0.151	0.560
Count R^2	0.835	0.871	0.917
observations	158	209	145

TABLE 2: SELLER DEVIATIONS FROM NARROW RATIONAL SELF-INTEREST: ACCEPTING OFFERS SHORT OF EQUILIBRIUM.

Notes. Significance levels are indicated by a : $p < .01$, b : $p < .05$, c : $p < .10$. Binary Logit models of decision: *accept* (1) vs. decision: *reject* (0) with standard errors clustered twoways at the individual and at the matching group level. *Full specifications including all six Kohlberg classes are reported in the paper's Online appendix E.3.*

DEPENDENT VARIABLE: SELLERS' DECISION TO ACCEPT (1)/REJECT (0) – OFFERS SHORT OF EQUILIBRIUM. INSTRUMENT: CON×POST.

$x_j \downarrow$ cluster/cluster mean $\rightarrow$	*2neg*$/-3.441$	*3neg*$/-1.209$	*5neg*$/-0.221$	*6neg*$/-0.665$
Intercept	−5.619	−11.562[c]	−0.023	0.298
	(5.256)	(6.484)	(1.903)	(2.893)
deviation	1.272[b]	−5.765[b]	−0.810	−0.022
	(0.619)	(2.463)	(1.366)	(1.358)
con×post· S(0,10,0)		2.650[a]		
		(0.804)		
con×post · S(0,10,10)		4.622[a]		
		(1.411)		
con×post	0.610[a]	−3.020[a]	0.187[b]	0.187[b]
	(0.194)	(0.664)	(0.084)	(0.110)
S(0,10,0): yes		−1.675[b]		
		(0.768)		
S(0,10,10): yes		−1.244		
		(1.238)		
Kohlberg class 1	−0.462	−0.209	0.112	−0.074
	(0.350)	(0.385)	(0.241)	(0.213)
Kohlberg class 2	−1.443[b]	−1.143	−0.101	−0.326
	(0.704)	(0.923)	(0.360)	(0.336)
Kohlberg class 3	0.641	0.081	0.233	0.231
	(0.586)	(0.561)	(0.176)	(0.243)
Kohlberg class 4	0.456	1.370[b]	0.099	0.183
	(0.437)	(0.547)	(0.337)	(0.299)
Kohlberg class 5	0.656	−0.347	−0.085	0.262
	(0.746)	(0.611)	(0.245)	(0.237)
Kohlberg class 6	−0.169	0.251	−0.251[a]	−0.306
	(0.300)	(0.483)	(0.211)	(0.242)
Period	−0.104[b]	−0.037	−0.035	−0.010
	(0.044)	(0.040)	(0.040)	(0.042)
Risk aversion	0.133	0.694[a]	0.046	−0.153[b]
	(0.217)	(0.110)	(0.101)	(0.069)
Age	−0.031	−0.103	−0.034	−0.034
	(0.134)	(0.133)	(0.048)	(0.062)
Gender:female	2.290[b]	0.944[c]	0.059	0.577[c]
	(1.020)	(0.568)	(0.337)	(0.343)
Extraversion	0.225[b]	0.255[a]	0.029	0.008
	(0.091)	(0.061)	(0.033)	(0.049)
Neuroticism	0.042	−0.066	−0.015	−0.004
	(0.036)	(0.046)	(0.026)	(0.030)
Psychoticism	0.123[c]	−0.139[c]	0.045	0.007
	(0.073)	(0.080)	(0.043)	(0.061)
Business/Economics	−3.288	2.867[c]	−0.754[c]	−0.622
Law	0.885	0.469	1.038[a]	0.536
Social and Behavioral Sciences	−0.570	−0.996	−0.363	0.473
Education	−1.245[c]	−0.693	−0.070	−0.752
Engeneering		0.314	[1.642]	−0.685
Medicine	−0.459	−3.606[a]	0.003	−0.715
Languages	1.087	1.072	0.131	0.761
observations/Count R^2	209/0.904	145/0.821	258/0.593	267/0.663

TABLE 3: SELLER DEVIATIONS FROM NARROW RATIONAL SELF-INTEREST. ACCEPTING OFFERS SHORT OF EQUILIBRIUM.

Notes. See notes to Table 2.

$S[0, 10, 10]$ with ample leeway to compensate. *Kohlberg class 5* is active for the two clusters which group offers shortest of equilibrium with an average deviation of -3.441 and -2.487 but not for offers closer to equilibrium. The second instrument *con×post* remains active up to the cluster of offers which fall least, i.e. by an average of 0.221, short of equilibrium. The two instruments therefore indicate that indeed, sellers require compensation for their lesser opportunities, a mechanism which counteracts the winner's curse. Turning to offers *above* equilibrium which sellers may deem include such a compensation, the effects reverse. Both instruments change sign, as do their interactions with $S[0, 10, 0]$ and $S[0, 10, 10]$.

An interesting question is why, for offers short of equilibrium, *con×post* should be active across a wider domain, and why therefore different amounts of compensation should be associated with each instrument. *Kohlberg class 5* measures how strongly an individual prefers social contract reasoning; it is foremost an affective[25] and arguably coarser instrument; *con×post* is an inverse cognitive measure of the extent to which an individual distinguishes between criteria of *Kohlberg classes 3,4* and *Kohlberg classes 5,6*, not unlike Georg Lind's C-score which captures individuals' *moral competence* by how much they can differentiate between all six *Kohlberg classes*, after having formed an opinion. A person with greater ability to discern the relevant ethical criteria, might also with greater ease perceive the relevance of *small* compensations for the violation of these criteria. The analysis on the seller side remains intricate since the degree to which sellers can deviate from equilibrium is censored by acquirers' offers, and the exact amount of resources available for compensation is unknown to them.

Note, however, that the main bone of contention between (Chlaß et al., 2019) and (Sugden and Wang, 2020) lies with *buyers*. Sugden and Wang (2020) report that inequality of opportunity entails destructive behaviour on *all* sides; I find that the party with greater freedom to choose becomes altruistic and compensates the party who finds herself with little room for manoeuvre. The same is observed in (Chlaß and Riener, 2015/2025) where parties transfer *all* payoff to the opponent with lesser opportunity, a compensation which we identify by two distinct causal methods: an experimental intervention, and, as a redundant check, the instrument *Kohlberg class 5*. Finally, a word on the nature of this reaction. Sugden and Wang (2020) argue that inequality of opportunity culminates in an emotional destruction of resources. What I observe here is much more in the spirit of a preference, rationally applied.

A sorrow with instruments is their exogeneity. To meet this concern, Appendix G shows in Figure A8 that subjects' *Kohlberg scores* are similarly distributed across roles and treatments such that, as advertised, there are no grounds to suspect a carry-over effect from the experimental task onto the M-J/C-T. Figure A9 shows that *Kohlberg scores* are very similarly distributed across all studies I conducted following up (Chlaß et al., 2019), looking at parties for whom we found a link between the instrument, and their actions in these respective experiments.[26] Then again, these scores might systematically differ compared to roles whose actions are *not* linked to the instrument. Thereof, only (Chlaß and Riener, 2015/2025) is ex ante balanced on gender in each session (a critical control for the instrument). For sixty-six statistical comparisons (Wilcoxon Rank Sum tests), six reject the null at the 5% level, roughly three more than should occur by chance, see online Appendix G.3.1. After controlling for fields of study in Online appendix G.3.2, none of these differences prevail: *Kohlberg class 5* no longer differs

[25]The idea in developmental psychology is that when individuals start to deem an ethical criterion *binding* in that it becomes an integral part of their motivation function, i.e. *internalize* that criterion, violations of that criterion start to cause emotional reactions: concern, rage, and – guilt (**?**, p. 9).

[26]These are proposers and responders in yes-no and ultimatum bargaining games from (Chlaß et al., 2019); buyers and sellers from this paper; proposers and responders in ultimatum and commitment games from (Chlaß and Miettinen, 2026), B players choosing the rules of the game from (Chlaß and Riener, 2015/2025), recipients in dictator games submitting hypothetical choices in (Chlaß and Moffatt, 2012/2017), and donors donating to a charitable cause with hidden prices (Chlaß et al., 2023), the latter conducted with a different M-J/C-T version in a *different* subject pool.

across treatments or roles. Note that two of these treatments give a player the option to take all opportunities from her opponent and feature substantial altruism – statistically explained by *Kohlberg class 5* –, see Appendix D, whereas players in the remaining four treatments follow their pure self-interest.[27] In Online appendix G, I also show that *Kohlberg class 5* and *con×post* neither differ across roles nor treatments in the current study, in (Chlaß and Miettinen, 2026), or in (Chlaß and Moffatt, 2012/2017), after controlling for critical conditions of the instrument. All studies vary the number of effective opportunities across roles and/or treatments and were conducted on the same subject pool.[28]

8 Conclusion

I point out errors in (Sugden and Wang, 2020). First, there is no treatment intervention. All options to turn cards are designed to induce the same expected outcome; an inequality in their number therefore does not consitute an inequality in opportunities since the authors define an opportunity as an option which produces a distinct outcome. Second, the game by which the authors measure individuals' responses to this 'treatment intervention', represents the actual treatment intervention: I show that the vendetta game used to measure subjects' emotions after experiencing 'unequal' opportunities, systematically privileges one player over the other in terms of opportunities. The party in question is the player who begins he game, i.e. the loser of the previous card game. I append an earlier paper with identical hypotheses but different results. The paper develops a game which systematically privileges one player over the other and shows by means of two instruments from (Chlaß et al., 2019) that the party who enjoys greater freedom to choose, compensates the party with lesser freedom of choice; the party with lesser freedom to choose requires such a compensation. The game is a common value auction and compensation is achieved through the common value. The errors I point out reconcile the differences in results between (Sugden and Wang, 2020) and (Chlaß et al., 2019,Chlaß and Riener, 2015/2025). Sugden and Wang (2020) find no attempts at compensating the 'unequal opportunities' they induce – which is not surprising, since there is no such inequality. Chlaß et al. (2019) and Chlaß and Riener (2015/25), however, find substantial attempts at compensation. Sugden and Wang (2020) report that all parties are upset by 'unequal opportunities', i.e. *both* players induce vendettas, rather than the privileged party becoming more altruistic. This is not surprising either: the winner from the card game has a sizeable disadvantage in *opportunities* in the vendetta game; the loser from the card game a sizeable disadvantage in *payoff* stemming from the card game.

Acknowledgements. I thank Christoph Göring for programming the zTree Code and his assistance in conducting the experiment. The experiment was funded by the Max Planck Society for the Advancement of Science.

Data Integrity. All treatments have been reported. Supplementary material including the online appendix is available from: https://www.chlass.de/research

[27]The regression for *con×post* in Online appendix G.3.3 shows that weak cognitive carry-over effects may potentially occur in (Chlaß and Riener, 2015/2025), cognitive, because *con×post* measures the correlation between *Kohlberg 3,4* and *Kohlberg 5,6* reasoning, potential, because they may also be randomization failures. *competitive payoffs×roleB: yes* has *p-value* =0.087; *competitive payoffs×treatment:lie, ref: spy* has *p-value* =0.046. These occur for treatment competitive payoffs where the opponent's effective opportunities can be taken away, compared to payoff neutrality (where this is not the case). Note, however, that these effects go into the wrong direction. Subjects in *competitive payoffs* score *higher* on the instrument which correlates *negatively* with preferences for equal rights; meaning that if the instrument shows an effect in some regression, its effect must be interpreted as a conservative estimate for the variable which is instrumented.

[28]Linear regressions do, in some cases, show significant coefficients for some *field of study* Dummy. Biserial correlation matrices in Online appendix F.1 do, however, not show any connection between the instruments and fields of study. The significant coefficients are therefore a result of typical multicollinearity between Dummies for *field of study* and the *constant* of the linear model.

A Courses of play in the simplified vendetta game

A.1 subtree 1: L:+0 at the outset

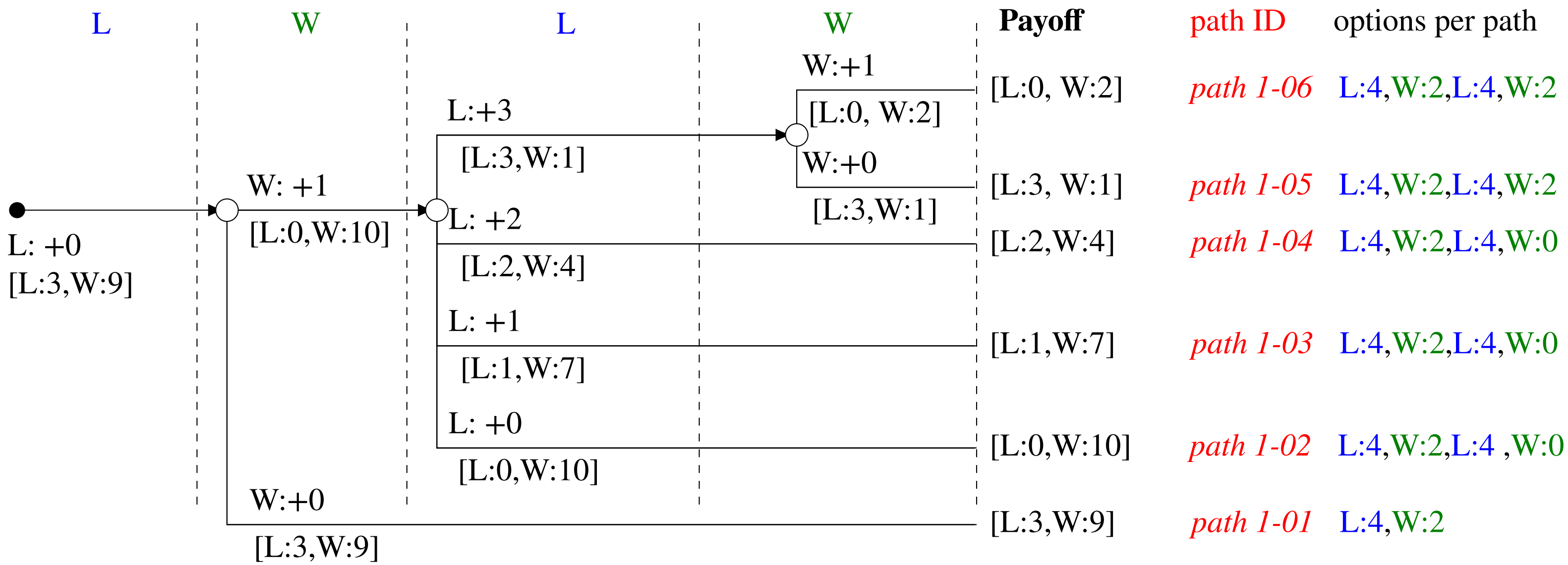


FIGURE A1: THE FIRST SUBTREE INDUCED BY L CHOOSING L=+0 AT THE BEGINNING OF THE SIMPLIFIED VENDETTA GAME.

Notes. In order to get a visual of the game and to illustrate that the vendetta game distributes opportunities unevenly, I first assume a simplified stopping rule according to which the game ends either if at two subsequent decision nodes, neither player has taken anything (i.e. each player needs to state only once, not twice that she does not want to take anything and wishes to end the game), or if the player next called upon to draw (but not necessarily both players) cannot take anything from her opponent anymore (has no options left). The full game with the stopping rule as implemented by Sugden and Wang (2020) is analyzed in Appendix I, Table A8. Table A8 uses essentially the same notation as the tree above. Above, L:+3, [L:3, W:1], for instance, denotes a subtree induced by $\mathscr{L}$ taking 9 tokens from $\mathscr{W}$ which adds three tokens (hence: L:+3) to her own; [L:3, W: 1] denotes the resulting distribution of tokens (the previous zero tokens for $\mathscr{L}$ plus three; the previous 10 tokens by $\mathscr{W}$ minus 9) at the beginning of the subtree. The first player called upon to play in this subtree is $\mathscr{W}$ who can take nothing, i.e. $\mathscr{W}$:+0, or take three tokens, i.e. $\mathscr{W}$:+1, from $\mathscr{L}$; the subtree therefore contains path IDs *1-05* and *1-06*. In Table A8, the same subtree is simply called: L: [L:3, W:1] since it is always the same, irrespectively of how much L has taken to open it.

A.2 subtree 2: L:+1 at the outset

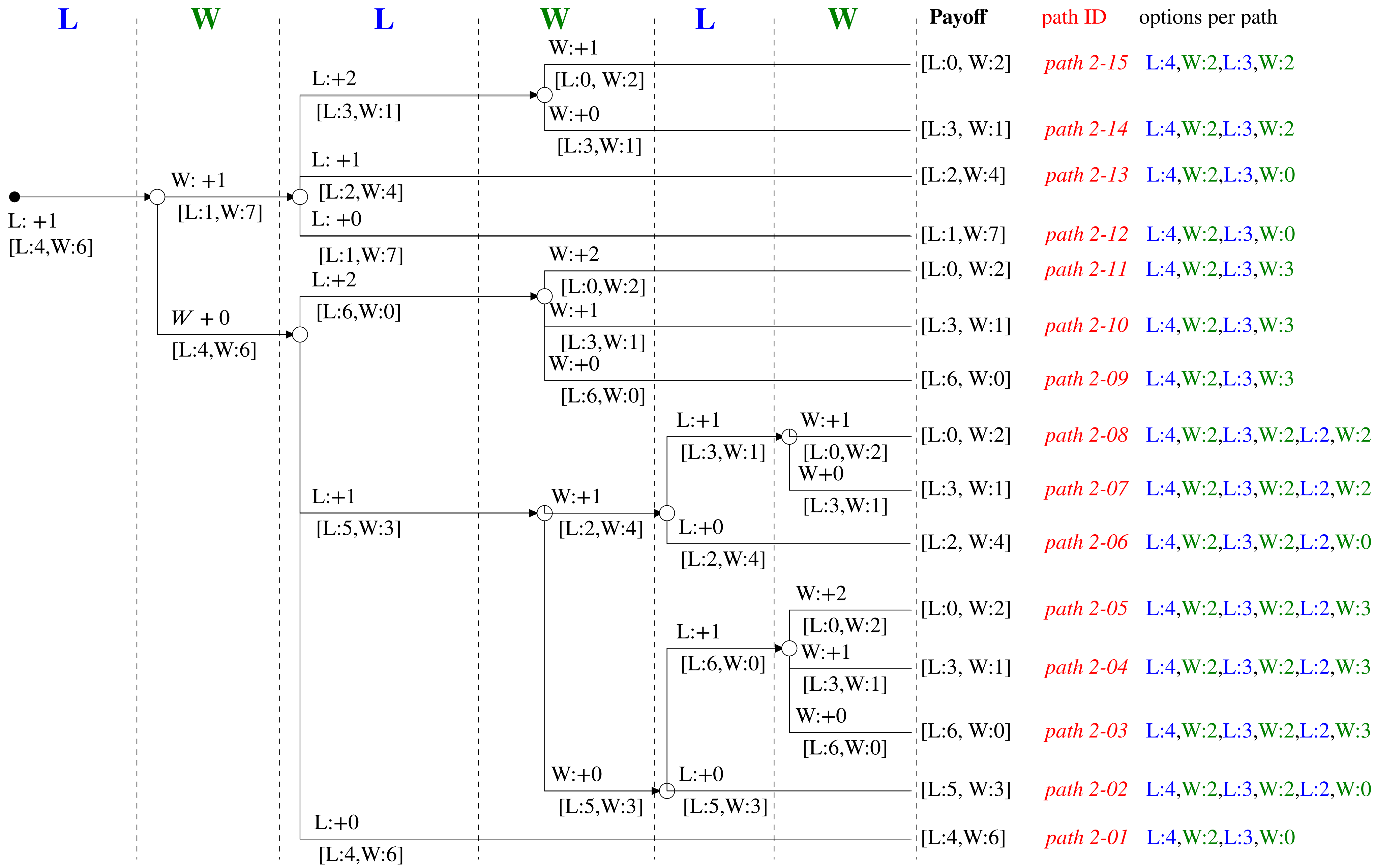


FIGURE A2: THE SECOND SUBTREE INDUCED BY L CHOOSING L:+1 AT THE BEGINNING OF THE SIMPLIFIED VENDETTA GAME.

A.3 subtree 3: L:+2 at the outset

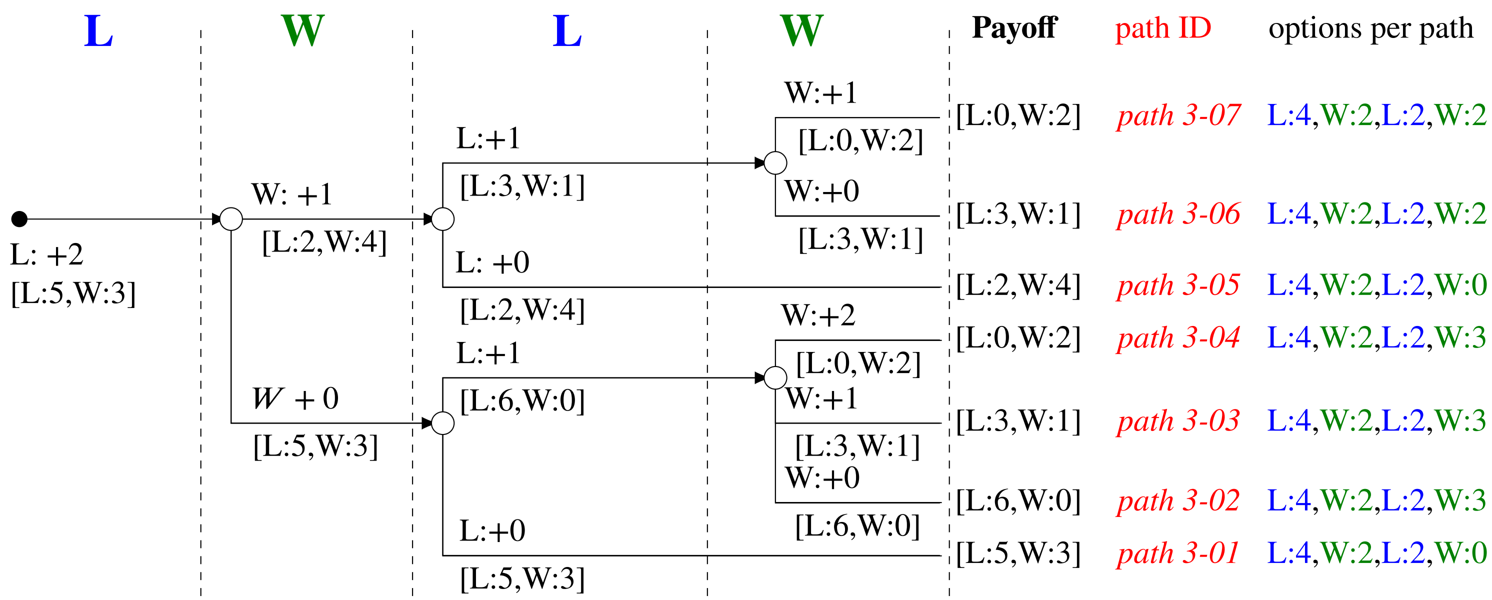


FIGURE A3: THE THIRD SUBTREE INDUCED BY L CHOOSING L:+2 AT THE BEGINNING OF THE SIMPLIFIED VENDETTA GAME.

A.4 subtree 4: L:+3 at the outset

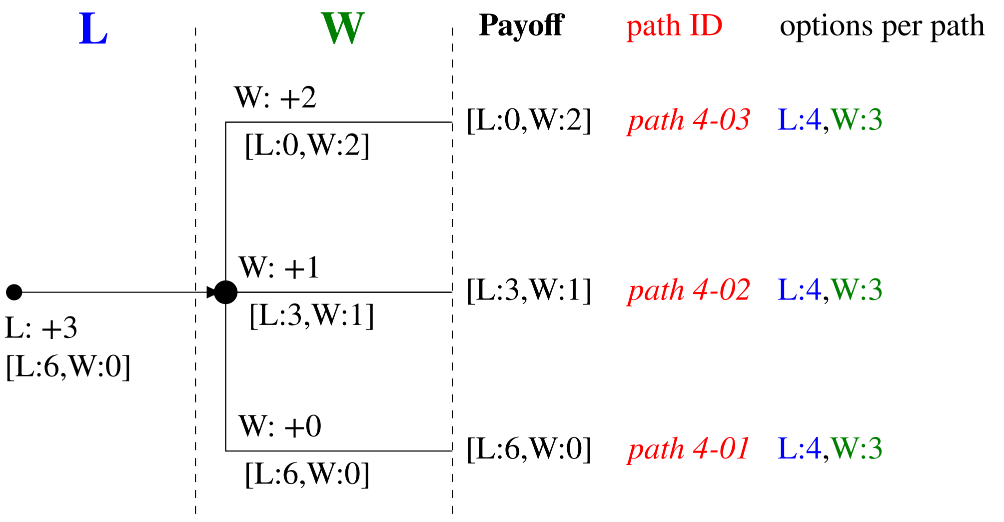


FIGURE A4: THE FOURTH SUBTREE INDUCED BY L CHOOSING L:+3 AT THE BEGINNING OF THE SIMPLIFIED VENDETTA GAME.

Notes. The fourth subtree illustrates the second part of the simplified stopping rule best: the game stops because the next player called upon to draw is $\mathscr{L}$. However, she cannot make any choice since she cannot take any block of three tokens from $\mathscr{W}$. $\mathscr{W}$ has at most two tokens to take. As mentioned earlier, the vendetta game in (Sugden and Wang, 2020) requires that *neither* player has any more block of three tokens to take and is analyzed in Table A8.

B Proofs

Corollary 1.

Proof. $f(v) = U(0, 10)$ has $E(v|v \geq p) = \frac{p+10}{2}$. A price equal to this conditional expected value has $p - q \cdot \frac{p+10}{2} = 0$ which yields $p = \frac{10q}{2-q}$. All prices below this threshold yield B zero payoff. □

Corollary 2.

Proof. A triangular distribution with lower bound 0, upper bound 10, and mode 0 has intercept $\frac{2}{10-0}$ and therefore $f(v) = 0.2 - 0.02v$. $E(v|v \geq p) = \frac{\int_p^{10} v(0.2-0.02v)\,dv}{\int_p^{10}(0.2-0.02v)\,dv}$ which is $\frac{\frac{2}{300}p^3 - \frac{1}{10}p^2 + \frac{10}{3}}{\frac{1}{100}p^2 - \frac{1}{5}p + 1}$. A price equal to q times this conditional expected value is $-10\frac{q}{-3+2q}$. □

Corollary 3.

Proof. A triangular distribution with lower bound 0, upper bound 10, and mode 10 has $f(v) = 0.02v$. $E(v|v \geq p) = \frac{\int_p^{10} v0.02v\,dv}{\int_p^{10} 0.02v\,dv}$ which is $\frac{\frac{20}{3} - \frac{2}{300}p^3}{1 - \frac{1}{100}p^2}$. A price equal to q times this conditional expected value is $\frac{1}{2}\frac{30-20q-10\sqrt{9+12q-12q^2}}{-3+2q}$. □

C Rules, payoffs, and the instrument in (Chlaß et al., 2019)

C.1 Two allocation procedures with payoffs to induce indifference

Fig. A5a) YES-NO GAME Fig. A5b) ULTIMATUM GAME

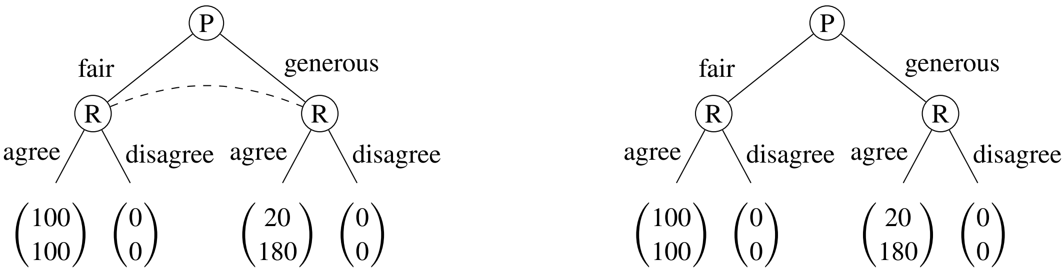


FIGURE A5: ILLUSTRATIONS OF THE TWO ALLOCATION PROCEDURES WHERE THE RIGHTS STRUCTURE WAS FOUND TO MATTER. **SOURCE:** (Chlaß et al., 2019, p. 111), condensed.

Notes. Indifferent proposers choose the same allocation in both games and believe that responders accept with certainty in both games. Indifferent responders always accept in each game and believe that proposers choose the same allocation with certainty in both games.

C.2 Count of effective opportunities for the procedures in Fig. A5

	A	D
F	(100,100)	(0,0)
G	(20,180)	(0,0)

FOR FIG. A5A): YES-NO GAME.

	AA	AD	DA	DD
F	(100,100)	(100,100)	(0,0)	(0,0)
G	(20,180)	(0,0)	(20,180)	(0,0)

FOR FIG. A5B): ULTIMATUM GAME.

Notes. F: fair split with payoffs (proposer, responder): (100,100), G: generous split with payoffs (20,180). A: agree, D: disagree. In YES-NO, proposer and responder each have two strategies (F and G; A and D). In ULTIMATUM, the proposer has two (F, G) strategies; the responder four: AA (agree to (100,100) and to (20,180), AD (agree to (100,100) but disagree to (20,180), DA: disagree to (100,100) but agree to (20,180), and DD: disagree to (100,100) and to (20,180). It is easy to see that each strategy yields distinct payoffs for some contingency. AA yields responder payoffs 100 and 180 which are distinct from AD with responder payoffs 100 and 0, from DA which has responder payoffs 0 and 180 and from DD which has 0 and 0. The count clearly includes the strictly dominated options DD, and analogously, also D in the yes-no game, for that matter. We postulate inequity aversion over effective opportunities, i.e. $u_i(s_i, s_j; b_i, b_j) - \beta_i max\{\#S_i - \#S_j, 0\} - \alpha_i max\{\#S_j - \#S_i, 0\}$ where player i's utility depends on the payoffs and beliefs (b) entailed by i's own and the opponent's pure strategies $s_{i,j}$ in sets $S_{i,j}$; i's dislike β_i of a disadvantage in the nr. (#) of effective opportunities, and her dislike α_i of an advantage in effective opportunities.

C.3 Indifferent subjects: choices over procedures link to Kohlberg 5 and con×post

preference →	YES-NO GAME		ULTIMATUM GAME	
role (P, R) →	proposer (P)	responder (R)	proposer (P)	responder (R)
nr. of obs.	42	21	24	21
pre	[−0.09]	[−0.09]	[−0.03]	[0.22]
con × post	−0.22 [a]	[−0.14]	[−0.02]	−0.69 [b]
Kohlberg 5	0.22 [a]	[−0.15]	0.31 [b]	0.64 [a]
simpler	[−0.01]	0.09 [a]	–	–
Count R^2	0.74	0.76	0.83	0.81

Significance levels of z-tests (referring to marginal effects as displayed) are indicated by a : $p < .01$, b : $p < .05$.

TABLE A1: PREFERENCES FOR THE YES-NO AND ULTIMATUM GAME LINK TO KOHLBERG 5, CON × POST, AND A SIMPLICITY RANKING. **SOURCE:** Chlaß et al., (2019), p.123, condensed.

Notes. *pre* denotes 'preconventional' moral argumentation (*Kohlberg classes 1 & 2*), con×post the interaction between preconventional moral argumentation (*Kohlberg 1 & 2*) and 'conventional' moral argumentation (*Kohlberg 3 & 4*), see Appendix H. [Square brackets] denote the marginal effect of an insignificant variable of theoretical interest, if added to the specification. Models were *tested down*, i.e. insignificant variables removed. There were altogether 38 indifferent responders. Since the yes-no game does not only distribute decision rights more equally but is also simpler than the ultimatum game, subjects were asked to rate the procedures in terms of simplicity in some sessions. Therefore, we only have complete data for 21 rather than all 38 indifferent responders.

C.4 More purely procedural aspects of the two procedures

purely procedural criterion	Yes-No game	Ultimatum game	ethical motivation (Kohlberg 5)
equality of decision rights	+	-	yes
equality of information rights	+	-	yes
transparency	-	+	yes
simplicity	+	-	no

TABLE A2: PURELY PROCEDURAL PROPERTIES OF THE TWO ALLOCATION PROCEDURES.
SOURCE: (Chlaß et al., 2019, p. 117), condensed.

C.5 Subjects who are not indifferent across the procedures in Fig. A5: choices between procedures and their link to Kohlberg 5.

role	cluster nr. (# nr. of obs.)	preferred game	material advantage *where?*	material advantage *size*	payment	Kohlberg 5
Proposers	1 & 2 (#31)	indifference (#10)	ultimatum	11.50	cannot pay	
		yes-no (#15)	ultimatum	1.00	9/15	24% $p = 0.00$
		ultimatum (#6)	ultimatum	14.17	2/6	29% $p = 0.00$
	3 (#17)	indifference (#6)	ultimatum	40	cannot pay	29% $p = 0.02$
		yes-no (#10)	yes-no	9	5/10	
		ultimatum (#1)	ultimatum	50	1/1	
Responders	1 (#22)	indifference (#7)	ultimatum	9.29	cannot pay	27% $p = 0.08$
		yes-no (#6)	ultimatum	98.33	3/6	46% $p = 0.01$
		ultimatum (#9)	ultimatum	26.67	1/9	
	2&3 (#33)	indifference (#12)	yes-no	32.08	cannot pay	31% $p = 0.01$
		yes-no (#9)	yes-no	22.78	6/9	
		ultimatum (#12)	yes-no	33.75	4/12	-22% $p = 0.04$

TABLE A3: SUBJECTS WHO DO NOT CHOOSE IDENTICAL STRATEGIES AND/OR DO NOT BELIEVE ALL OPPONENTS CHOOSE THE SAME STRATEGY: CHOICES AND THEIR RELATION TO *Kohlberg class 5*.

Notes. Subjects were matched by allocation choices and beliefs in all procedures via Ward's clustering (Ward, 1963). In merged clusters 1&2, there are n=31 proposers. 15 prefer the yes-no game, 9 of which pay for the game. On average, these proposers expect a 1 ECU (3 Eurocents) advantage in the ultimatum game. A one point increase in Kohlberg class 5 increases the likelihood of preferring the yes-no game by 24%, *p-value* = 0.00, over being indifferent (avg. marginal effect taken over all individuals). Kohlberg class 5 also explains proposers preferring the ultimatum game over being indifferent. These proposers expect an advantage of 14.17 ECU (42.5 Eurocents) in the ultimatum game and yet, also act against their self-interest since they mostly don't pay; payment costs 15 Eurocents if one is drawn; accounting for the 50% chance to be drawn, payment costs, on expectation, 7.50 Eurocents which is less than the expected advantage. In cluster 3, Kohlberg class 5 explains (29%, *p-value* =0.02)[29] why subjects do not pursue their expected advantage in the ultimatum game (proposers who state to be indifferent when they expect an advantage of 40 ECU in the ultimatum game) compared to subjects who do pursue their advantage (preference for the yes-no game when they expect a material advantage in that game). In responder cluster 1, Kohlberg 5 explains subjects preferring the yes-no game (46%, *p-value* = 0.01) and also subjects stating indifference (27%, *p-value* = 0.08) when they expect a (sizeable) advantage in the ultimatum game. In clusters 2&3, Kohlberg 5 explains subjects stating indifference when they expect an advantage in the yes-no game (31%, *p-value* =0.01) compared to subjects who pursue that advantage in the yes-no game. Subjects who state indifference also make more use of Kohlberg 5 (22%, *p-value* =0.04) if compared to subjects who state a preference for the ultimatum game. All preference groups which act against their self-interest, can be explained by Kohlberg 5.

[29] Note that the original paper codes the yes-no game as 1 (success), and indifference as 0 (failure) and therefore reports the reverse (negative) sign, see (Chlaß et al., 2019, Online appendix G, p. 20)

D Rules, payoffs, and the instrument in (Chlass and Riener, 2015/25)

D.1 Equal Rights versus three asymmetric procedures with identical payoffs

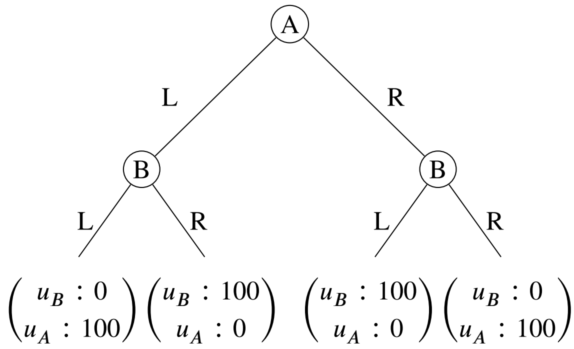
Fig. A6-i) EQUAL RIGHTS

B \ A	L	R
L	100 / 0	0 / 100
R	0 / 100	100 / 0

Fig. A6-ii) FABRICATION/SABOTAGE

B \ A	L	R
LL^A	100 / 0	100 / 0
RL^A	0 / 100	0 / 100
LR^A	0 / 100	0 / 100
RR^A	100 / 0	100 / 0

Fig. A6-iii) SPYING



Notes. Figs. A6-i) & A6-iii): L: Left; R: Right; Fig. A6-ii): LL^A: B chooses L herself, and replaces (or fabricates) A's choice by L^A. Procedure(s) A6-ii)-FABRICATION and A6-ii)-SABOTAGE share the same normal form. In SABOTAGE, B overrides A's (unknown) choice, in FABRICATION, B transmits A's (unknown) choice and her own choice to an uninformed Dummy player who must implement the transmitted choices. In A6-iii)-SPYING, B looks up A's choice before she makes her own choice, whereby A does not know how B chooses. [Player indexes are kept the same across normal and extensive form such that player 2, i.e. A, moves first, and player 1, i.e. B, moves last in the extensive form; payoffs at the terminal nodes are listed by increasing player index – first B, then A.]

FIGURE A6: B EITHER OPTS INTO EQUAL RIGHTS, I.E. FIG. A6-I); OR REDUCES A'S DECISION RIGHTS IN FIG. A6-II); OR INCREASES HER OWN DECISION AND INFORMATION RIGHTS, FIG. A6-III).

D.2 Count of effective opportunities and information rights

A \ Player B	L	R
L	(100,0)	(0,100)
R	(0,100)	(100,0)

FIG. A6-i)

A \ Player B	LL^A	RL^A	LR^A	RR^A
L	(100,0)	(0,100)	(0,100)	(100,0)
R	(100,0)	(0,100)	(0,100)	(100,0)

FIG. A6-ii): FABRICATION/SABOTAGE.

A \ Player B	LL	LR	RL	RR
L	(100,0)	(100,0)	(0,100)	(0,100)
R	(0,100)	(100,0)	(0,100)	(100,0)

FIG. A6-iii): SPYING.

Notes. In A6-i) EQUAL RIGHTS, L yields A payoffs 100 and 0; R yields A 0 and 100, and A therefore has two effective opportunities; the same holds for player B. Both players have the same information rights: neither knows how the other moves; each distinguishes 2 pairs of terminal nodes. In A6-ii), L and R yield A the same payoffs and she has zero effective opportunities. For B, LL^A and RR^A yield identical payoffs each of 0 and 0; LR^A and RL^A yield her each 100 and 100. B therefore has two effective opportunities. Neither player knows the opponent's choice. In A6-iii), L yields A payoffs 100,100,0,0; R yields A payoffs 0,100,0,100, and A therefore has two effective opportunities. B has four effective opportunities since all strategies yield her distinct payoffs. In A6-iii), B has the advantage in information: she distinguishes all four terminal nodes of the game.

D.3 Experimental setup and results.

Experiment. **Design**. (Only) B chooses the procedure, the costless default being a 50-50 draw, i.e. the toss of a fair coin: B can either make Fig. A6-i) more likely, or, depending on the treatment, Fig. A6-ii) or Fig. A6-iii). In all treatments, a (narrowly) rational self-interested B makes Fig. A6-ii) or Fig. A6-iii) certain, takes the entire pie of 100 ECU (5 Euros with 1€ $\widehat{=}$ US\$1.28 at the time) and leaves 0 ECU for A. **Results**. With FABRICATION, 9 of 44 Bs pay for EQUAL RIGHTS, and 17 of 25 who arrive in Fig. A6-ii) give all payoff to A. In SPY, 5 of 54 Bs pay for Fig. A6-i) and *nobody* in Fig. A6-iii) gives all payoff to A. In SABOTAGE, 2 of 54 Bs pay for EQUAL RIGHTS, and 20 of 28 Bs in Fig. A6-ii) give all payoff to A. In separate treatments, Bs' expectation as to how much A will punish B's choice symbolically (i.e. punishment is too weak to induce reciprocity) shows that Bs are most selfish (SPY) where they expect punishment to be highest. In three further treatments (payoff neutrality), payoffs are changed such that B cannot change the distribution of decision rights. Virtually no departures from rational self-interest occur in these treatments.

D.4 Choices of Equal Rights and altruism under unequal rights both link to Kohlberg 5

Dependent Variable: three Departures from Narrow Rational Self-Interest (Base Category)			
x_j ↓ specification →	(1)	(2)	(3)
Kohlberg class 1	-0.071^c (0.039)	-0.079^b (0.039)	
Kohlberg class 2	−0.036 (0.045)		
Kohlberg class 3	0.086^b (0.040)	0.088^b (0.037)	
Kohlberg class 4	0.041 (0.053)		
Kohlberg class 5	0.105^b (0.042)	0.101^a (0.037)	
Kohlberg class 6	−0.031 (0.041)		
Dummy LIE (FABRICATION)	0.444^a (0.084)	0.438^a (0.086)	
Dummy SABOTAGE	0.240^a (0.062)	0.221^a (0.055)	
Risk aversion	−0.004 (0.019)	−0.002 (0.018)	−0.021 (0.022)
Envy	0.054 (0.063)	0.055 (0.061)	−0.015 (0.073)
Age	−0.003 (0.011)	−0.002 (0.011)	0.007 (0.014)
Gender:female	−0.093 (0.066)	−0.090 (0.066)	−0.085 (0.081)
Business/Economics	0.036	0.057	−0.044
Medicine	-0.374^b	-0.374^b	-0.435^a
Law	0.139	0.167	0.208
Social and Behavioral Sciences	−0.073	−0.059	0.058
IT	-0.740^a	-0.713^a	-0.754^a
Languages	-0.412^a	-0.420^a	-0.340^b
Education	−0.071	0.075	−0.026
Nagelkerke's R^2	0.60	0.58	0.16
B participants	150	151	151

Table A4: B's procedural and allocation choices in one multinomial model (marginal effects displayed). **Source:** (Chlass and Riener, 2015/25, Appendix Table A10).

Notes. Multinomial Logit regression of 4 nominal categories l where $l = 0 \leftrightarrow$ rational self-interest with B participants opting into Fig. A6-ii) or Fig. A6-iii) and taking all payoff; $l = 1 \leftrightarrow B$ participants who toss a fair coin, arrive in A6-ii) or A6-iii) and take all payoff; $l = 2 \leftrightarrow B$ participants who opt into A6-i) and arrive there, and $l = 3 \leftrightarrow B$ participants who opt into A6-ii) or A6-iii) and give all payoff, B participants who opt into A6-ii) or A6-iii) but end up in A6-i), and B participants who toss a fair coin, arrive in A6-ii) or A6-iii) and give all payoff. Results do not depend upon whether Bs who opt into A6-ii) or A6-iii) and end up in A6-i), are classified as $l = 3$ or not. Marginal effects represent the average over all individual marginal effects. We report the *negative* average marginal effect of each variable x_j on the base category $l = 0$, revealing by how much a one unit change in the variable increases the likelihood of a departure from rational self-interest. Standard errors of marginal effects are obtained by the Delta method. Specifications include (1) a full (inefficient) model with all variables and controls, (2) an efficient model including significant variables only but full controls, and (3) a biased model with controls only to showcase their explanatory power. One B participant mistakenly jumped all Kohlberg 4 items, resulting in 150 observations for specification (1) only. Results are fully robust to other specifications, i.e. treating the categories independently in (less efficient) simple binary logits where they reach a 1% significance level even on full specifications, see appendices O.1 and O.2 in (Chlaß and Riener, 2015/2025).

E Human and robot treatments: comparison

E.1 Offers and MAOs for the Robot treatment

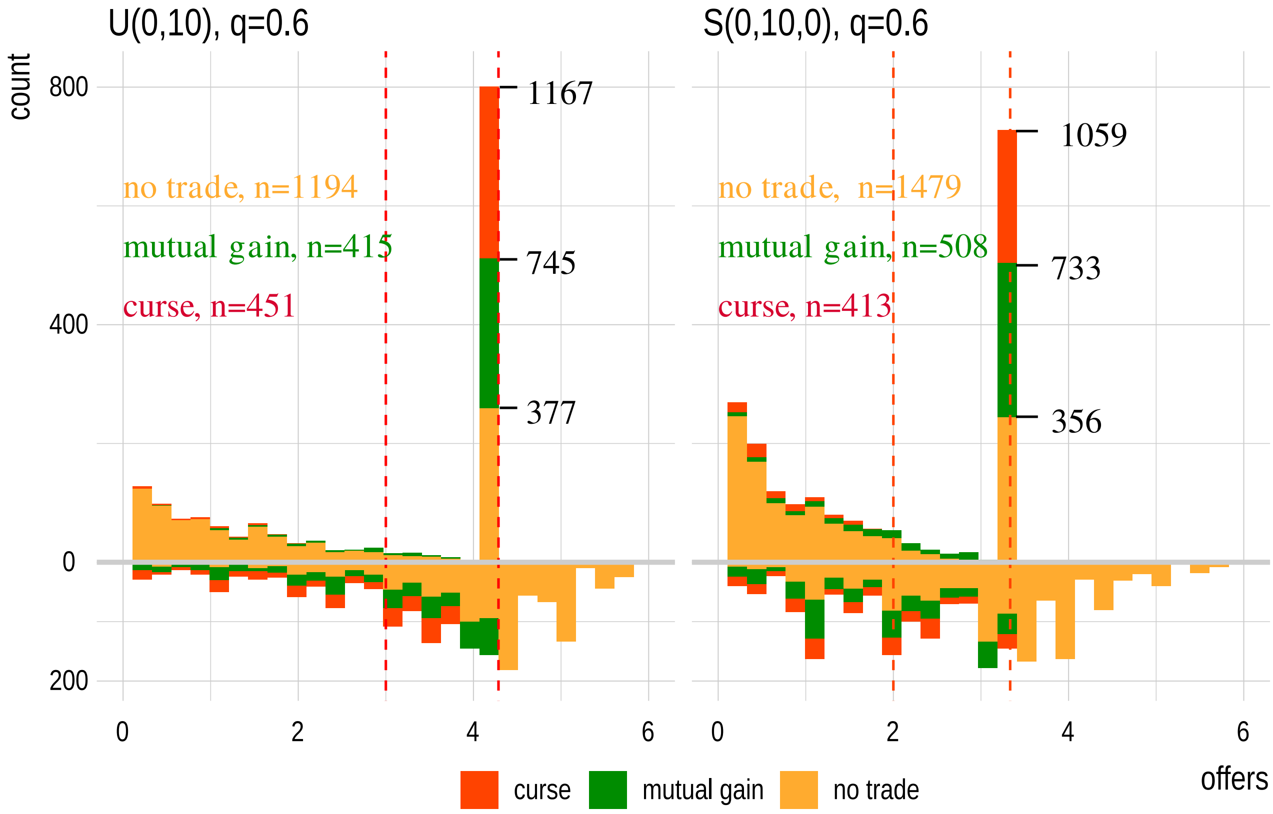


Figure A7: Mirror histograms of robot offers and MAOs by human players; all rounds.

Notes. Upward histogram: Perfect Bayesian Nash equilibrium offers made by robot buyers. Downward histogram: minimally acceptable offers (MAOs) submitted by human sellers. Left: treatment $U[0,10]$; Right: treatment $S[0,10,0]$. Humans submit their MAOs by means of the strategy method and are randomly matched with a specific robot; there is, therefore, no correspondence between an upward pointing bar of offers, and a downward pointing bar of MAOs. Yellow: offers which did not exceed the seller's MAO/MAOs which did not fall below the robot's offer; Green: offers and MAOs which led to an actual gain for robot and human; Red: offers which led to an actual loss for the human player. *To better visualize the data, the one single large bar on top has been resized proportionally to the other bars; a new legend just for this single bar has been added directly to the single large bar. The red and green shaded areas of Figure A7 can directly be compared to those in Figure 2 of the main text.*

E.2 Robot-human versus human-human treatments: key figures

Treatments $U[0,10]$, q=0.6 and $S[0,10,0]$, q=0.6 were conducted both with robot buyers and human sellers (see Figure A7 in Appendix E.1 above), and with human buyers and human sellers (see Figure 2, main text).

	human		robot	
	$S[0,10,0]$	$U[0,10]$	$S[0,10,0]$	$U[0,10]$
losses	0.08 (40)	0.07 (34)	0.17 (413)	0.22 (451)
rejected	0.44 (211)	0.44 (208)	0.62 (1479)	0.58 (1194)
mutual gain	0.48 (229)	0.49 (228)	0.21 (508)	0.20 (415)
nr of interactions	480	470	2400	2060

Table A5: Key figures for the robot-human, and for the human-human treatments.

F Information on seller clusters

Note. The main text presents clusters in the same order (cluster nr.) as returned by the WARD procedure. **This table lists them by the size of the deviation**, to see how the percentage of rejection decreases.

cluster nr.↓ (text)	deviation	mean	std. dev.	nr of obs.	reject	reject per cent	$S[0, 10, 0]$ (share)	$U[0, 10]$ (share)	$S[0, 10, 10]$ (share)
2neg	$[-4.57, -2.87]$	-3.441	0.441	209	176	84.21%	0.41	0.34	0.25
1neg	$[-2.85, -2.23]$	-2.487	0.204	158	116	73.42%	0.36	0.46	0.18
4neg	$[-2.21, -1.41]$	-1.794	0.219	188	119	63.30%	0.19	0.53	0.28
3neg	$[-1.39, -0.96]$	-1.209	0.135	145	86	59.31%	0.50	0.30	0.20
6neg	$[-0.94, -0.44]$	-0.665	0.131	267	172	64.42%	0.35	0.41	0.24
5neg	$[-0.43, -0.01]$	-0.221	0.117	258	131	50.78%	0.32	0.40	0.28
1pos	$[0.01, 1.08]$	0.556	0.288	624	194	31.09%	0.32	0.35	0.32
2pos	$[1.09, 2.07]$	1.555	0.267	535	36	7.21%	0.31	0.23	0.46
3pos	$[2.08, 5.89]$	2.876	0.664	414	18	4.35%	0.31	0.23	0.46

Table A6: Range, average, and standard error of potential seller deviations from equilibrium (price offer minus equilibrium offer), and rejection rates for all clusters obtained. The table panel to the right shows how many deviations in a cluster belong to each treatment.

G Kohlberg class 5: exogeneity

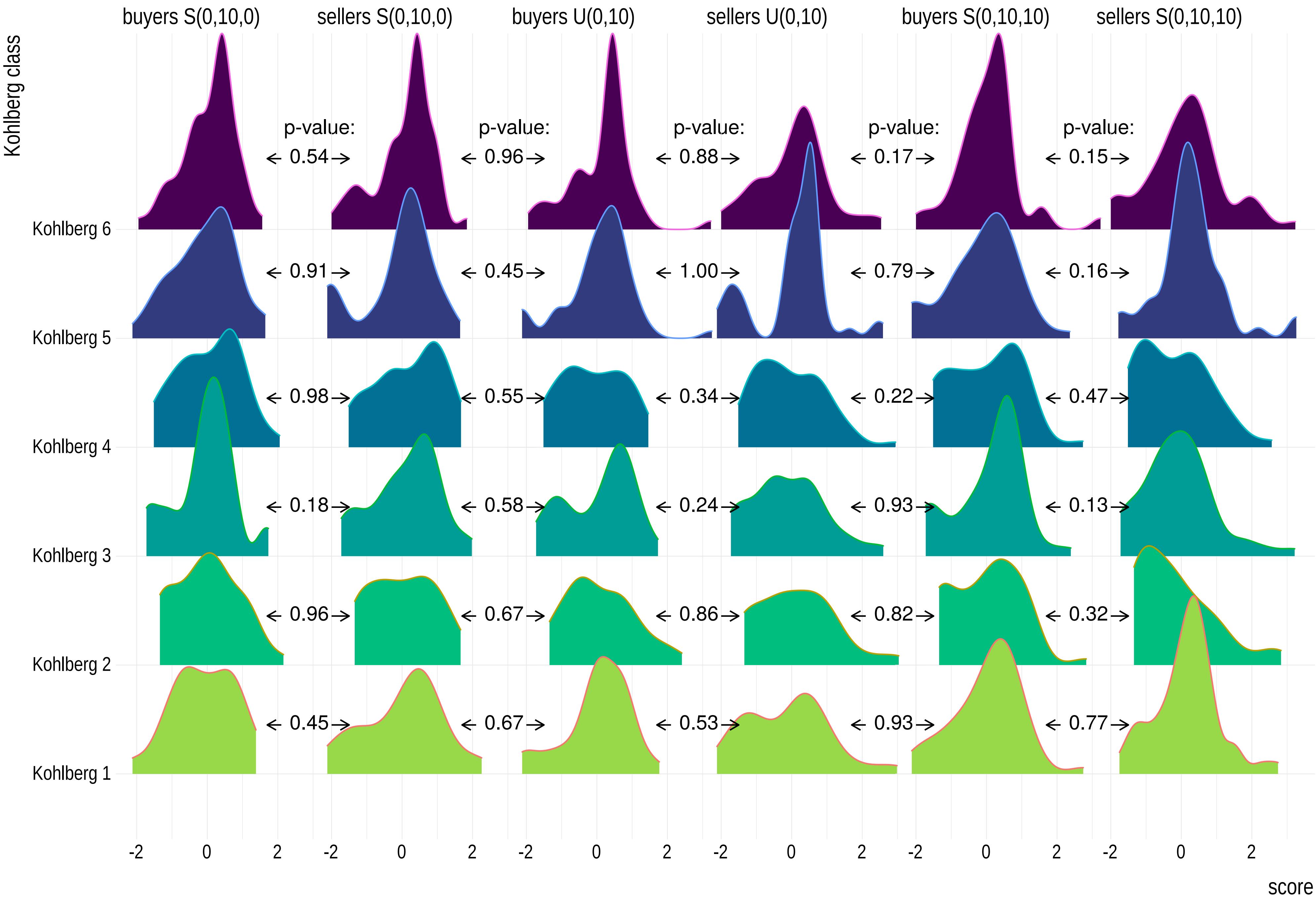


FIGURE A8: DISTRIBUTIONS OF KOHLBERG SCORES DO NOT DIFFER ACROSS ROLES AND TREATMENTS.

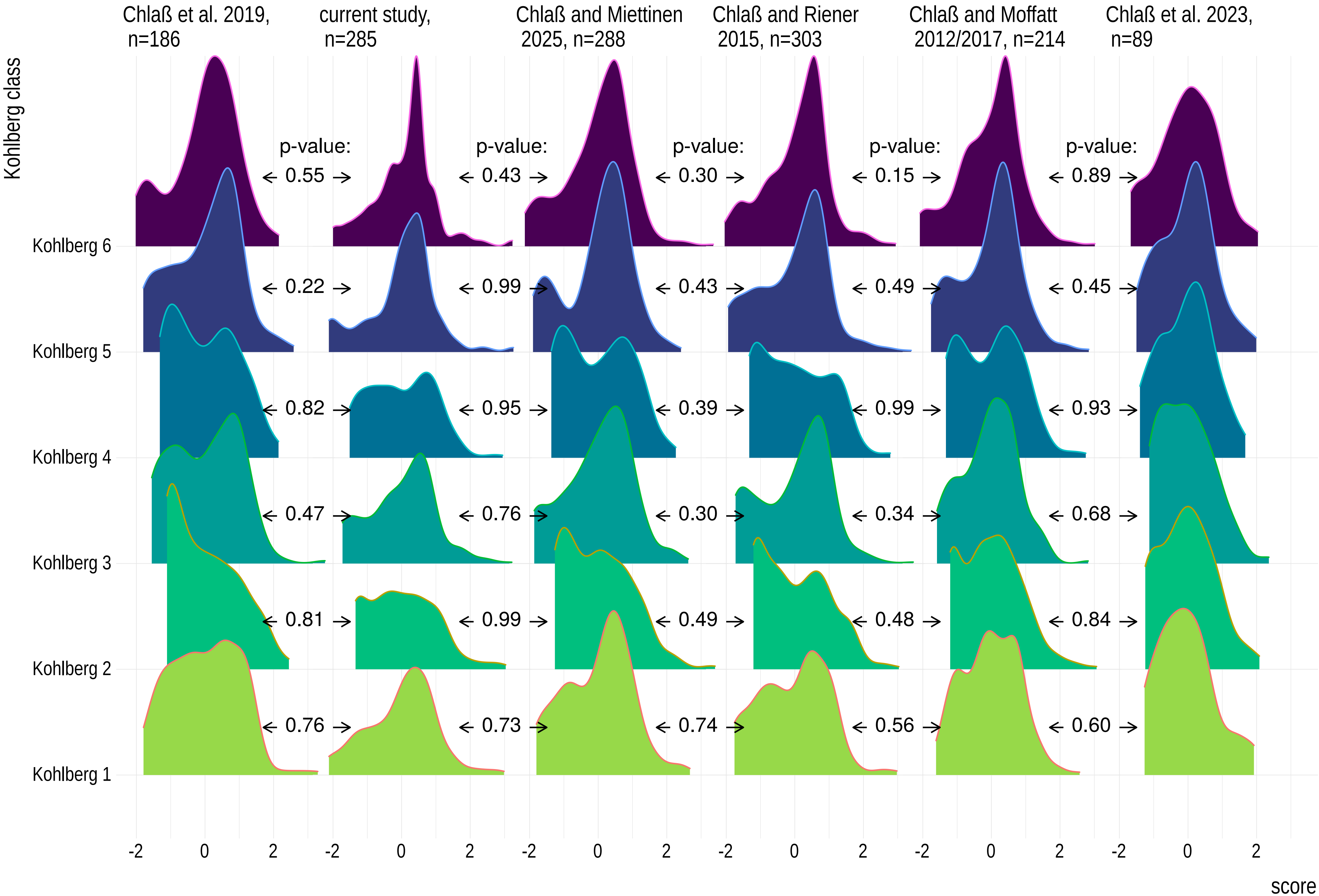


FIGURE A9: DISTRIBUTIONS OF KOHLBERG SCORES DO NOT DIFFER ACROSS STUDIES.

H Kohlberg's taxonomy of moral argumentation and its correspondence with a row of social preference models

Six categories of Lawrence Kohlberg's taxonomy of moral argumentation and a list of economic preference models built on the respective criteria listed in each category.

preconventional argumentation	preference models
Kohlberg 1. Orientation toward punishment and obedience, physical and material power. Rules are obeyed to avoid punishment. ***Kohlberg 2.*** Naïve hedonistic orientation. The individual conforms to obtain rewards.	(...)
conventional argumentation	preference models
Kohlberg 3. Orientation toward inter-individual mutual relations. "Good boy/girl" orientation to win the approval and maintain the expectations of one's immediate group. The individual conforms with norms and expectations and shows good intentions to avoid disapproval. One earns approval by being "nice".	guilt aversion (Battigalli and Dufwenberg 2007), inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), social norms (Bicchieri, 2006; Krupka & Weber, 2013), reciprocal preferences (Falk and Fischbacher 2006, Dufwenberg and Kirchsteiger 2007), preferences for equal expected payoffs (Bolton et al. 2005); preferences for kind procedures (Sebald 2010); image (Ellingsen & Johannesson, 2008; Andreoni & Bernheim, 2009)
Kohlberg 4. Orientation toward law and order, and in particular societal expectations and moral rules from outside one's immediate peer group since these maintain and ensure the continuity of the social order.	rule-following
postconventional argumentation	preference models
Kohlberg 5. Orientation toward the social contract. Duties are defined by the social contract and the equality of rights stipulated therein. Emphasis is on mutual commitment and obligation in a liberal democratic basic order.	Chlaß et al.'s (2019) purely procedural preferences: equality of decision rights and information rights;
Kohlberg 6. Orientation toward universal ethical principles of conscience such as Kant's categorical imperative. Rightness of an act is derived from abstract and consistent ethical principles such as the inalienability of human rights, the free will, and individuals' freedom to choose. Ethical principles are a priori truths inherent in rational beings as laid down by Kant's categorical imperative.	Alger and Weibull's (2013) Homo Moralis

Notes. Source: (Chlaß and Miettinen, 2026, p. 20) and the references therein. Economic preference models listed are *preferences for kind procedures.* Sebald, A. (2010), Attribution and Reciprocity, Games and Economic Behavior, 68, pp. 339-352; *preferences for equal expected payoffs.* Bolton, G., Brandts, J.,Ockenfels A. (2005), Fair Procedures: Evidence From Games Involving Lotteries, Economic Journal, 115, pp. 1054-1076; *guilt aversion.* Battigalli, P. and M. Dufwenberg (2007), Guilt in Games, American Economic Review, Papers and Proceedings, 97, pp. 170-176; *reciprocal preferences.* Falk, A., Fischbacher, U. (2006), A Theory of Reciprocity, Games and Economic Behavior, 54, pp. 293-315, Dufwenberg, M., Kirchsteiger, G. (2004), A Theory of Sequential Reciprocity, Games and Economic Behavior, 47, pp. 268-98; *inequity aversion.* Bolton, G., Ockenfels A. (2000) ERC - A Theory of Equity, Reciprocity and Competition, American Economic Review, 90, pp. 166-193, Fehr, E., Schmidt, G. (1999), A Theory of Fairness, Competition and Cooperation, Quarterly Journal of Economics, 114, pp. 817-868; *social norms.* Bicchieri, C. (2006). The Grammar of Society, Cambridge, UK: Cambridge University Press; Krupka, E. L., and Weber, R. A. (2013), Identifying social norms using coordination games: Why does dictator game sharing vary?, Journal of the European Economic Association, 11(3), pp. 495-524; *image.* Ellingsen, T. and Johannesson. M. (2008), Pride and Prejudice: The Human Side of Incentive Theory, American Economic Review, 98 (3), pp. 990-1008, Andreoni, J. and Bernheim, B.D. (2009), Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects, Econometrica, 77(5), pp. 1607-1636; *purely procedural preferences*. Chlaß N., Güth, W., Miettinen, T. (2019), Purely Procedural Preferences – Beyond Procedural Equity and Reciprocity, European Journal of Political Economy, 59, pp. 108-128; *Homo Moralis.* Alger, I. and Weibull, J.W. (2013), Homo Moralis - Preference Evolution Under Incomplete Information and Assortative Matching, Econometrica 81(6), pp. 2269-2302.

SELLERS' DECISIONS TO ACCEPT/REJECT IN CLUSTERS 1POS (BINARY LOGIT) 2POS AND 3POS (POWER LOGITS)– OFFERS ABOVE THE EQUILIBRIUM.

x_j ↓ (cluster nr.)/mean→	*1pos*/0.55/624	*2pos*/ 1.56/535	*3pos*/ 2.88/414
Intercept	1.945 (2.318)	19.796^a (4.449)	22.438^b (9.819)
deviation	1.343^a (0.376)	0.027 (0.961)	−0.093 (0.533)
Kohlberg class 5 · S(0,10,0)		2.445^b (1.190)	3.364^a (1.300)
Kohlberg class 5 · S(0,10,10)		1.868^b (0.945)	−0.315 (1.919)
con·post · S(0,10,0)		−4.324 (3.030)	1.848 (1.731)
con·post · S(0,10,10)		−5.517^c (3.078)	0.499 (0.856)
Kohlberg class 5	0.212 (0.144)	−1.172^c (0.917)	−1.150 (1.901)
con · post	−0.282^b (0.137)	5.502^c (3.134)	−0.078 (0.707)
S(0,10,0): yes		−0.352 (1.946)	1.400 (1.358)
S(0,10,10): yes		−0.871 (1.534)	0.372 (1.041)
Kohlberg class 1	0.218 (0.223)	−0.264 (0.409)	0.620 (0.477)
Kohlberg class 2	0.049 (0.308)	0.295 (0.345)	−0.700 (0.671)
Kohlberg class 3	0.097 (0.211)	0.046 (0.363)	−0.308 (0.394)
Kohlberg class 4	−0.281 (0.293)	−0.162 (0.317)	0.155 (0.464)
Kohlberg class 6	−0.174 (0.129)	0.525 (0.383)	1.164^b (0.460)
Period	0.010 (0.020)	−0.030 (0.030)	0.037 (0.032)
Risk aversion	0.054 (0.092)	0.276^c (0.153)	0.482^b (0.215)
Age	−0.148^a (0.051)	−0.109 (0.111)	−0.195 (0.200)
Gender:female	0.132 (0.414)	−0.008 (0.549)	−1.329 (1.025)
Extraversion	0.020 (0.022)	0.062^b (0.031)	0.009 (0.082)
Neuroticism	0.028 (0.026)	−0.096^a (0.030)	0.012 (0.033)
Psychoticism	0.011 (0.051)	−0.055 (0.074)	−0.151 (0.111)
Business/Economics	0.463	−0.391	0.722
Law	−0.102	−2.032	−0.881
Social and Behavioral Sciences	0.607	−0.332	−1.504
Education	−0.337	0.250	−1.343
Engeneering	1.047	1.442	−2.702
IT	0.800	0.731	−1.960
Languages	0.435	0.941	0.838

TABLE A7: SELLERS DEVIATIONS: ACCEPTING OFFERS ABOVE EQUILIBRIUM.

Notes. For cluster *1pos*: binary Logit. For clusters *2pos* and *3pos*: Power Logits. Successes (acceptance) recoded as failures (rejection) and vice versa, coefficients of of power logits reported with reversed sign such that they can be interpreted as coefficients for *acceptance*. Errors clustered at the individual and the matching group level.

I The vendetta game without simplified stopping rule

path ID↓	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$
1-01	4	2	4	2														
1-02	4	2	4	2	4	0	4	0										
1-03	4	2	4	2	4	0	4	0	3	0	3							
1-04	4	2	4	2	4	0	4	0	3	0	3	0	2	0	2			
1-05	4	2	4	2	4	0	4	0	3	0	3	0	2	0	2	2	0	2
1-06	4	2	4	2	4	0	4	0	3	0	3	0	2	0	2	2	0	2
1-07	4	2	4	2	4	0	4	0	3	0	3	0	2	0	2	2		
1-08	4	2	4	2	4	0	4	0	3	0	3	0	2	2	0	2		
1-09	4	2	4	2	4	0	4	0	3	0	3	0	2	2	0	2		
1-10	4	2	4	2	4	0	4	0	3	0	3	0	2	2				
1-11	4	2	4	2	4	0	4	0	3	0	3	2	0	2				
1-12	4	2	4	2	4	0	4	0	3	0	3	2	0	2				
1-13	4	2	4	2	4	0	4	0	3	0	3	2						
1-14	4	2	4	2	4	0	4	0	3	0	2	0	2					
1-15	4	2	4	2	4	0	4	0	3	0	2	0	2	2	0	2		
1-16	4	2	4	2	4	0	4	0	3	0	2	0	2	2	0	2		
1-17	4	2	4	2	4	0	4	0	3	0	2	0	2	2				
1-18	4	2	4	2	4	0	4	0	3	0	2	2	0	2				
1-19	4	2	4	2	4	0	4	0	3	0	2	2	0	2				
1-20	4	2	4	2	4	0	4	0	3	0	2	2						
1-21	4	2	4	2	4	0	4	0	3	2	0	2						
1-22	4	2	4	2	4	0	4	0	3	2	0	2						
1-23	4	2	4	2	4	0	4	0	3	2								
1-24	4	2	4	2	4	0	4	L: [L:2,W:4]										
⋮																		
1-30																		
1-31	4	2	4	2	4	0	4	L: [L:3,W:1]										
⋮																		
1-33																		
1-34	4	2	4	2	4	L: [L:1,W:7]												
⋮																		
1-54																		
1-55	4	2	4	2	4	L: [L:2,W:4]												
⋮																		
1-61																		
1-62	4	2	4	2	4	L: [L:3,W:1]												
⋮																		
1-64																		
1-65	4	2	4	2	3	2	3											
1-66	4	2	4	2	3	2	3	2	2	2	2							
1-67	4	2	4	2	3	2	3	2	2	2	2	2	0	2				
1-68	4	2	4	2	3	2	3	2	2	2	2	2	0	2				
1-69	4	2	4	2	3	2	3	2	2	2	2	2						
1-70	4	2	4	2	3	2	3	2	2	2	2	0	2					
1-71	4	2	4	2	3	2	3	2	2	2	2	0	2					
1-72	4	2	4	2	3	2	3	2	2	2	2							
1-73	4	2	4	2	3	2	3	2	2	2	0	2						
1-74	4	2	4	2	3	2	3	2	2	2	0	2						
1-75	4	2	4	2	3	2	3	2	2	2								
1-76	4	2	4	2	3	2	3	2	2	0	2							
1-77	4	2	4	2	3	2	3	2	2	0	2							
1-78	4	2	4	2	3	2	3	2	2									
1-79	4	2	4	2	3	2	3	3	0	3								
1-80	4	2	4	2	3	2	3	3	0	3	0	2	0	2				
1-81	4	2	4	2	3	2	3	3	0	3	0	2	0	2				
1-82	4	2	4	2	3	2	3	3	0	3	0	2						
1-83	4	2	4	2	3	2	3	3	0	3								
1-84	4	2	4	2	3	2	3	3	0	2	0	2						
1-85	4	2	4	2	3	2	3	3	0	2	0	2						
1-86	4	2	4	2	3	2	3	3	0	2								
1-87	4	2	4	2	3	2	3	3										

path ID↓	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$
1-88	4	2	4	2	3	2	3	0	3									
1-89	4	2	4	2	3	2	3	0	3	L: [L:2,W:4]								
⋮																		
1-95																		
1-96	4	2	4	2	3	2	3	0	3	L: [L:3,W:1]								
⋮																		
1-98																		
1-99	4	2	4	2	3	2	3	L: [L:2,W:4]										
⋮																		
1-105																		
1-106	4	2	4	2	3	2	3	L: [L:3,W:1]										
⋮																		
1-108																		
1-109	4	2	4	2	3	2	2	2	2									
1-110	4	2	4	2	3	2	2	2	2	L: [L:6,W:0]								
⋮																		
1-118																		
1-119	4	2	4	2	3	2	2	2	2	0	2							
1-120	4	2	4	2	3	2	2	2	2	0	2	2	0	2				
1-121	4	2	4	2	3	2	2	2	2	0	2	2	0	2				
1-122	4	2	4	2	3	2	2	2	2	0	2	2						
1-123	4	2	4	2	3	2	2	2	2	L: [L:3,W:1]								
⋮																		
1-125																		
1-126	4	2	4	2	3	2	2	L: [L:6,W:0]										
⋮																		
1-134																		
1-135	4	2	4	2	3	2	2	0	2									
1-136	4	2	4	2	3	2	2	0	2	2	0	2						
1-137	4	2	4	2	3	2	2	0	2	2	0	2						
1-138	4	2	4	2	3	2	2	0	2	2								
1-139	4	2	4	2	3	2	2	L: [L:3,W:1]										
⋮																		
1-141																		
1-142	4	2	4	2	3	L: [L:6,W:0]												
⋮																		
1-150																		
1-151	4	2	4	2	W: [L:1,W:7]													
⋮																		
1-171																		
1-172	4	2	4	L: [L:5,W:3]														
⋮																		
1-204																		
1-205	4	2	4	L: [L:6,W:0]														
⋮																		
1-213																		
1-214	4	2	4	0	4													
1-215	4	2	4	0	4	L: [L:1,W:7]												
⋮																		
1-225																		
1-226	4	2	4	0	4	L: [L:2,W:4]												
1-232																		
1-233	4	2	4	0	4	L: [L:3,W:1]												
⋮																		
1-235																		
1-236	4	2	4	L: [L:1,W:7]														
⋮																		
1-256																		
1-257	4	2	4	L: [L:2,W:4]														
⋮																		
1-263																		
1-264	4	2	4	L: [L:3,W:1]														
⋮																		
1-266																		

path ID↓	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$	$\mathscr{L}$	$\mathscr{W}$
2-01 ⋮ 2-107	4	L: [L:4,W:6]																
3-01 ⋮ 3-33	4	L: [L:5,W:3]																
4-01 ⋮ 4-09	4	L: [L:6,W:0]																

TABLE A8: PATHS OF PLAY IN THE FULL VENDETTA GAME.

Notes. Lists all paths of play in the vendetta game which ends if (i) in four consecutive rounds, no player took anything, or alternatively, (ii) *neither* player possessed any more three tokens to take. Options to take (that is, actions available at a given decision node) are sorted in ascending order and listed left to right (leftmost action: not taking anything, rightmost action: taking everything). Direction of movement through the game tree while denoting paths: top to bottom, left to right. Game tree is denoted as in Figs. A1, A2, A3, A4, rotated clockwise by 90°, i.e. flows top-to-bottom. Column header $\mathscr{L}$ or $\mathscr{W}$ in Table A8 denotes the player called upon to play for every decision node along a path of play. Note that, against extensive form convention, decision nodes where a player had no options to choose from, are still displayed as decision nodes if the game continues beyond that node (i.e. unless this happens at the end of a path). Each table cell counts the number of options a player can choose from when called upon to play. ▇ highlights subsequent decision nodes where $\mathscr{L}$ has lesser options to choose from than $\mathscr{W}$. ▇ denotes $\mathscr{L}$'s last exit option to prevent this before ▇; that is, her last chance to force $\mathscr{W}$ onto a different path of play such that $\mathscr{L}$ has again always more than or equally many options to choose from as $\mathscr{W}$. *Notations of subtrees.* To handle the vast number of paths, I work with subtrees. Table A8 first marks all paths opened by a subtree in its specific color. Take ▇. These are 21 paths of play, opened by subtree L:[L:1,W:7] which opens with a decision node where $\mathscr{W}$ has no options to choose from since she cannot take anything. Take L:[L:2,W:4]. The subtree opens 7 paths, beginning with $\mathscr{W}$ having zero options to choose from. Overall, we see that instances where $\mathscr{W}$ has more options to choose from than $\mathscr{L}$ are few and far between. In fact, they accumulate in subtree L:[L:6,W:0]. $\mathscr{L}$ can always avoid that subtree right before it opens.

References

Achen, C.H.(2002), Toward a new Political Methodology: Microfoundations and ART, Annual Review of Political Science, 4, pp. 423-450.

Bebeau, M. (2002) The Defining Issues Test and Four Component Model: Contributions to Professional Education, 31(3), Journal of Moral Education, pp. 271-295.

Charness, G., Levin, D. (2009), The Origin of the Winner's Curse: A Laboratory Study, American Economic Journal: Microeconomics, 1(1), pp. 207-236.

Chlaß, N. (2010), The Impact of Procedural Asymmetry in Games of Imperfect Information, Papers of the German Economic Association Annual Meeting 2010, available at https://hdl.handle.net/10419/37253.

Chlaß, N. (2011), On Individual Cursedness – How Personality Shapes Individuals' Sensitivity to Incur a Winner's Curse, Jena Economic Research Paper, #2011-027.

Chlaß, N., Gangadharan, L., Jones, K. (2023), Charitable Giving and Intermediation: A Principal-Agent Problem with Hidden Prices, Oxford Economic Papers, 75 (4), pp. 941-961.

Chlaß, N., Güth, W., Miettinen, T. (2019), Purely Procedural Preferences – Beyond Procedural Equity and Reciprocity, European Journal of Political Economy, 59, pp. 108-128.

Chlaß, N., Moffatt, P. G. (2012/2017), Giving in Dictator Games – Experimenter Demand Effect or Preference over the Rules of the Game? Jena Economic Research Paper #2012-044, University of Jena; CBESS Discussion Paper #17-05, University of East Anglia.

Chlaß N., Miettinen, T. (2026), Commitment in Sequential Bargaining – An Experiment, Journal of Economic Behavior and Organization, 242, art. no 107360.

Chlaß N., Riener, G. (2015/2025), Lying, Spying, Sabotaging – Procedures and Consequences –/Preferences for Fair Competition, Jena Economic Research Paper # 2015-016. Available at https://www.chlass.de/research.html

Eyster, E., Rabin, M. (2005), Cursed Equilibrium, Econometrica, 73, pp. 1623-1672.

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics, 10, pp. 171-178.

Greiner, B. (2015), Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE, Journal of the Economic Science Association, 1, pp. 114-125.

Kohlberg, L., Hersh, R. H. (1977/2009), Moral Development: A Review of the Theory, Theory into Practice, 16 (2), pp. 53–59.

Lind, G. (1977/2021), The Moral Competence Test. Available at: http://moralcompetence.net.

Lind, G. (2008), The Meaning and Measurement of Moral Judgment Competence Revisited - A Dual Aspect Model, In: D. Fasko, W. Willis (Eds), Contemporary Philosophical and Psychological Perspectives on Moral Development and Education. NJ, Hampton Press, Cresskill, pp. 185-220.

Pattanaik, P. K., Xu, Y. (2000), On Ranking Opportunity Sets in Economic Environments, Journal of Economic Theory, 93, pp. 48-71.

Piaget, J. (1948),The Moral Judgment of the Child, Free Press, Glencoe, Illinois.

Rest, J. (1979), Development in Judging Moral Issues, University of Minnesota Press, Minneapolis, MN.

Jones, P., Sugden, R. (1982), Evaluating Choice, International Review of Law and Economics, 2, pp. 47-65.

Sugden, R., Wang, M. (2020), Equality of Opportunity and the Acceptability of Outcome Inequality, European Economic Review, 130, art. no. 103597.

Trautmann, S. T. (2023), Procedural Fairness and Equality of Opportunity, Journal of Economic Surveys, 37, pp. 1697-1714.

Ward, J. H. (1963), Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 1963, 58, pp. 236-244.

IMPRESSUM
Jena Economics Research Papers
ISSN 1864-7057
Friedrich Schiller University Jena
Faculty of Economics and Business Administration
Carl-Zeiss-Str. 3
D-07743 Jena, Germany

Email: office.jerp@uni-jena.de
Editor: Silke Übelmesser
Website: www.wiwi.uni-jena.de/en/jerp