Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Fakultät Wirtschafts- und Sozialwissenschaften

Reciprocity Models Revisited: Intention Factors and Reference Values

Janna Hinz
Andreas Nicklisch

WiSo-HH Working Paper Series
Working Paper No. 25
April 2015

WiSo-HH Working Paper Series
Working Paper No. 25
April 2015

Reciprocity Models Revisited: Intention Factors and Reference Values

Janna Hinz, Universität Hamburg
Andreas Nicklisch, Universität Hamburg

ISSN 2196-8128

Font used: „TheSans UHH“ / LucasFonts

Kontakt:
WiSo-Forschungslabor
Von-Melle-Park 5
20146 Hamburg
E-Mail: experiments@wiso.uni-hamburg.de
Web: http://www.wiso.uni-hamburg.de/forschung/forschungslabor/home/

Reciprocity Models Revisited: Intention Factors and Reference Values*

Janna Hinz§ Andreas Nicklisch‡

April 29, 2015

Abstract

We present a test of the two most established reciprocity models, an intention factor model and a reference value model. We test characteristic elements of each model in a series of twelve mini-ultimatum games. Results from online experiments with nearly 500 subjects show major differences between actual behavior and predictions of both models: the distance of actual offers to the proposed reference value provides a poor measure for the kindness of offers, while a pairwise comparison of offers as suggested by the intention factor model cannot explain behavior in richer settings. We discuss possible combinations of both models describing our observations.

Keywords: *Experiments, Intentions, Mini-ultimatum Game, Reference Value, Reciprocity Models*
JEL-Classification: C52, C72, C91

*We gratefully acknowledge the many useful comments provided by Anke Gerber, Manuel Schubert, Irenaeus Wolff, and the participants of the 2014 GfeW meeting in Passau. This research is partly funded by the German Research Foundation (DFG), project Ni 1610/1-1.

§Department of Economics, University of Hamburg, von-Melle-Park 5, 20146 Hamburg, Germany, e-mail: janna.hinz@wiso.uni-hamburg.de

‡*Corresponding author:* Department of Economics, University of Hamburg, and German Research Foundation (DFG), Research Unit "Needs based Justice and Distribution Procedures", von-Melle-Park 5, 20146 Hamburg, Germany, e-mail: andreas.nicklisch@wiso.uni-hamburg.de

1 Introduction

Reciprocity is one fundamental cornerstone of human behavior, and an integral element for other-regarding preferences. The importance of reciprocal behavior for human interactions has been stressed by a large body of economic literature (e.g., Cox & Deck, 2005, Andreoni et al., 2003, Falk et al., 2003, Gächter & Thöni, 2007).[1] Consequently, behavioral economists have been striving to explain how people digress from self-interested behavior to reward kind actions and punish unkind actions of their opponents. Modeling reciprocity, however, has turned out to be a very complex endeavor. The specific formulation of reciprocal preferences follows predominantly two distinct ways: Rabin, 1993, and Dufwenberg & Kirchsteiger, 2004 focus on an intra-personal comparison according to reference values, whereas Falk & Fischbacher, 2006 rely on an inter-personal comparison using an intention factor to capture underlying motivations. This ambiguity has led subsequent studies to rely on one or the other approach (e.g., Ambrus & Pathak, 2011, Stanca et al., 2009). In this current study we test characteristic features of both theories in a number of mini-ultimatum games similar to the one used in the tradition of Bolton and Zwick (1995) and Falk et al. (2003). Particularly, we focus on two key differences between the two approaches: firstly, the reference value approach measures the extent of (un)kindness according the distance of the specific offer to the reference value, while the intention factor approach measures the unkindness by the inequity of the specific offer; secondly, the reference value approach assesses (un)kindness globally (i.e., considering all potential alternatives of the game), whereas the intention factor approach performs pairwise comparisons (i.e., one kind alternative can turn all other alternatives inevitable into fully intentional unkind alternatives). We show that both approaches have advantages and disadvantages when explaining actual decisions so that a combination of both approaches seems to provide a good description of behavior.

The general idea of reciprocal preferences is perhaps best summarized by the Latin principle 'quid pro quo.' The overarching non-parametric model by Cox et al. (2008) formalizes these words in the following way: suppose, a player ('Chip') has a number of alternatives from which he can choose one. His choice has consequences in terms of payoffs ('berries') not only for himself, but also for another player ('Chap'). Chap considers Chip's choice *blue* to be more generous than *red* if *blue* yields more berries to Chap than the choice of *red*, while Chip's gain from choosing *blue* and not *red* is at most as large as Chap's gain from choosing *blue* and not *red*. Reciprocity in this context means that the more generous Chip's choice, the more money Chap is willing to spend to increase

[1] Alternative approaches such as distributional concerns (e.g., Fehr & Schmidt, 1999) or guilt-aversion (e.g., Battigalli & Dufwenberg, 2007) have been shown to explain pro-social behavior partly, but not comprehensively (e.g., Andreoni et al., 2003).

Chip's earnings. Chap gains *immaterial* utility from Chip's material payoffs if Chip behaves generously, whereas Chap's immaterial utility can even be negative, if Chip choose a mean alternative. Consequently, Chap may want to punish Chip, for instance by passing the berries on to the birds.

Contemporary reciprocity models translate the idea of quid pro quo into well-defined closed preference models. We refer to the first approach as the "reference value model", first formalized by Rabin (1993).[2] In his belief-dependent model, reciprocity is analyzed for two-players, normal-form games. Dufwenberg & Kirchsteiger (2004; herafter D&K) extended Rabin's model of belief-dependent preferences to extensive n-player games. In both models, Chap ranks Chip's alternatives from the one yielding the least berries for *Chap* to the most berries for Chap. Half way between Chap's lowest and highest payoff lies Chap's *equitable payoff* dividing Chip's alternatives into unkind ones below the equitable payoff and kind ones above (hereafter, we denote the equitable payoff as the reference value). *Chip's (un)kindness towards Chap* increases in the difference between the payoff corresponding to Chip's choice and the reference value. We would like to stress that the reference value model measures the action's kindness by a "global assessment". That is, the midpoint of the entire set of alternatives in the game determines the reference value which, in turn, determines the kindness of a specific offer.

We test the predictive success of the reference value model according to two characteristics. Firstly, we check for the continuity of the reference value: we analyze whether the likelihood that Chap responds reciprocally increases in the distance between the equitable payoff and the payoff "normally resulting" from Chip's chosen action.[3] Secondly, we test for the predictive success of the reference value: varying the game, but keeping the reference value and the distance from the reference value constant, we analyze whether the likelihood that Chap responds unkindly remains constant.

The second class of reciprocity models, "intention factor models," contrast the reference value models in two ways: they decompose Chip's (un)kindness towards Chap into the intention term and the outcome term (e.g., Falk & Fischbacher, 2006, hereafter F&F). The first term determines whether Chap perceives Chip's action as intended or not, the second term determines the severeness of Chap's perceived (un)kindness. F&F place in Chap's immaterial partial utility from reciprocity prior importance on the difference of payoffs between Chip and Chap within an option before the difference to the other possible payoffs for Chap is considered. Notice that the intention factor model assesses the action's kindness by a pairwise comparison. Therefore, having one very kind action in the set of al-

[2]Other models incorporating reciprocity follow the same logic, but apply slightly different techniques (e.g., Cox et al, 2007).

[3]We clarify the meaning of "normally" below.

ternatives turns all other actions in the game to be unkind. This important – but often neglected – difference from reference value models has serious consequences for predictions in games with several alternatives.

The crucial importance of intentions for reciprocal behavior has been shown elsewhere (e.g. Falk et al., 2008). We test the predictive success of intentions by checking for the continuity of the intention factor: we analyze whether the likelihood that Chap responds unkindly increases, if the intention factor increases, keeping the inequity of Chip and Chap's payoffs constant. Secondly, we test for consistency of the intention factor: varying the game, but keeping the intention factor and the inequity of payoffs constant, we analyze whether the likelihood that Chap responds unkindly remains constant. In other words, we test whether an assessment of kindness relying on the pairwise comparison describes Chip's kindness properly.

For our purpose, we propose a series of twelve mini-ultimatum games. Some of them offer Chip two alternatives to choose from, some of them offer four alternatives. All of them allow Chap to reject a proposed alternative and forgo his own income for the sake of punishing Chip. The games are designed such that they allow us to asses the predictive success of reference value models and intention factor models. We retrieve our data in online experiments with almost 500 participants. As such, our analysis follows Sobel's (2005) criticism that existing reciprocity models seem to be fitting for specific situations, but lack a clear characterization of this very situation. Along the same line of arguments, there are some other studies discussing and testing the predictive success of reciprocity models. Firstly, they provide evidence of the importance of intentions for reciprocation: if there is no alternative but to behave unkindly, subjects reciprocate less severely (Falk et al., 2003); the same holds true if an action is taken that is not unambiguously kind, but selfish to some degree (Stanca et al., 2009). Secondly, Dhaene & Bouckaert (2010) elicit first and second order beliefs of participants in a sequential prisoners' dilemma and a mini-ultimatum game. They show that beliefs and behavior, particularly of second movers, are very consistent with D&K's reciprocity model. Furthermore, Pelligra (2011) varies systematically the outside options in a trust game, where the first mover's trusting is either kind or unkind for the second mover. Contrasting the theoretical predictions of the D&K model, the trustworthiness of the second mover remains constant across treatment conditions suggesting that other motives dominate behavior in this setting (cf., Pelligra, 2011). Finally, Nicklisch & Wolff (2012) test an overall characteristic of reference value models and intention factor models: if punishment is sufficiently cheap, reciprocation is modeled as a "all-or-nothing" decision. That is, if Chip behaves kindly (unkindly), Chap maximizes his utility by choosing the most kind (unkind) response possible. By means of a modified ultimatum game, the authors show that decisions for a majority of participants in a laboratory experiments do not follow this assumption.

Along this partly pessimistic assessment of contemporary reciprocity models, our data shows important shortcomings for both models when predicting behavior. Particularly, the continuity of the reference value model fails to characterize actual behavior: increasing the distance between the equitable payoff and the payoff of the actual offer does not necessarily correspond with increasing rejection rates. Moreover, variation of the game yields differences in the rejection rates although the reference value and the distance from the reference value remain constant. We conclude from those findings that the distance to the reverence value serves as a poor descriptor for the extent of (un)kindness. On the other hand, experimental results for simple games with two alternatives are nicely predicted by the intention factor model. The likelihood of rejection increases for increasing intention factors. However, there is little consistency between predictions and decisions in the richer games with four alternatives. We conclude from this that the pairwise comparison of alternative does not characterize behavior adequately, and suggest a combination of both approaches. This combination includes a global assessment for the intention of a choice and the inequity of an alternative for the extent of (un)kindness.

The remainder of this article is organized as follows: The following section (re)acquaints with both reciprocity models to be tested with an emphasis on the element we scrutinize. In Section 3 we introduce our experimental design and procedure. Section 4 presents results. In Section 5 we discuss our findings and suggest potential developments for reciprocity models reflecting our results. Section 6 concludes.

2 Reciprocity based utility

2.1 Sequential reciprocity according to Dufwenberg & Kirchsteiger

In their reference value model, Dufwenberg & Kirchsteiger (2004, D&K) provide a solution concept for belief-dependent reciprocal behavior in sequential n-player games they call "sequential reciprocity equilibrium". Their core premise is that extensive form games require an updating of players' beliefs as the play unfolds, making it necessary for Rabin's (1993) work to be extended and refined. More specifically, Chap's immaterial payoff component evolves dynamically with each definite action of Chip: as Chap's beliefs about Chip's future behavior need to be updated and possibly revised, so does the perceived kindness of Chip's actions.

Formally, let Chip be player j and Chap player i; j chooses an action a_j from his set of alternatives A_j. Suppose a_j affects Chip's and Chap's payoffs (π_j and π_i, respectively). Chap observes Chip's action. Then D&K define Chap's utility

function as follows:

$$U_i^{D\&K} = \pi_i(a_i, a_j) + \Upsilon_i \kappa_{ij}^{D\&K} \lambda_{iji}^{D\&K} \tag{1}$$

Utility consists of a material payoff π_i and an immaterial payoff component. The material part of U_i refers to the payoff assigned to the end note of a specific choice. The immaterial part of utility is initiated with an individual sensitivity parameter to reciprocal concerns, Υ_i. If $\Upsilon_i = 0$, then utility will equal material payoff, such as suggested by narrow self-interest. $\kappa_{ij}^{D\&K}$ denotes Chap's perceived (un)kindness of Chip's action, and $\lambda_{iji}^{D\&K}$ is the (un)kindness of Chap's response given Chap's belief of Chip's expectations concerning Chap's behavior in the consecutive game.

According to D&K, Chap's immaterial partial utility from reciprocity is proportional to the product of Chap's (un)kindness towards Chip by deviating from "the normally resulting way", $\lambda_{iji}^{D\&K}$, and Chip's (un)kindness towards Chap, $\kappa_{ij}^{D\&K}$. Notice that "the normally resulting way" means within this context that the game terminates in an end note which corresponds with Chip's expectation. Thus, both terms depend on Chap's second order belief concerning Chip's assumptions on how Chap proceeds the game. In our setting, this belief simplifies dramatically: we consider mini-ultimatum games where players in Chap's role can either accept or reject an offer (rejections result in zero payoff both for Chip and Chap). Therefore, for every alternative Chip chooses Chap can assume acceptance as his second order belief. Otherwise, Chip chooses inefficient (i.e., Pareto dominated) strategies which contradicts a general assumption of D&K's approach.

Chap's reference value separating Chip's actions into kind and unkind actions is the value half way between the lowest and highest payoff (in terms of berries) at the time when Chip makes his decision. More generally, i's equitable payoff $\pi_i^{e_j}$ (i.e., i's mean payoff following j's action) is:

$$\pi_i^{e_j} = 0.5 \max(\Pi_i') + 0.5 \min(\Pi_i'), \tag{2}$$

where Π_i' is the set of payoffs induced by j's efficient strategies. In turn, not considered are j's inefficient strategies, that is, strategies for which one finds a Pareto improvement – in terms of i's and j's payoffs – among j's strategies for any strategy choice of i (compare D&K, pp. 275-276).

Chip's (un)kindness towards Chap is then increasing in the difference between the payoff corresponding to Chip's kind (unkind) choice and the reference value:

$$\kappa_{ij}^{D\&K} = \pi_i - \pi_i^{e_j} \tag{3}$$

After Chap observes Chip move, it is on him to respond, again influencing the monetary outcomes for both players. That is, ranking Chap's alternatives from the one yielding the least berries to the most berries for *Chip*, the midpoint of the ranking determines Chap's reference value for the (un)kindness of his response. In other words, Chap's (un)kindness towards Chip is measured according to the distance between Chip's actual payoff to Chip's equitable payoff:[4]

$$\lambda_{iji}^{D\&K} = \pi_j - \pi_j^{e_i} \tag{4}$$

D&K's explicit quantification of (un)kindness with the equitable payoff as a reference value allows us to test the predictive success of their model:

$\text{Hyp}_{D\&K}$: *Decreasing the distance to the equitable payoff across actions implies non-increasing rejection rates for those actions, while keeping the distance constant implies a constant rejection rate.*

2.2 Intention-based reciprocity according to Falk & Fischbacher

Similarly to D&K, Falk & Fischbacher's (2006, F&F) approach differentiates between perceived (un)kindness and the (un)kindness of the response. Therefore, we can define Chap's utility function according to F&F also as:

$$U_i^{F\&F} = \pi_i(a_i, a_j) + \Upsilon_i \kappa_{ij}^{F\&F} \lambda_{iji}^{F\&F} \tag{5}$$

Within the context of our simple mini-ultimatum games, one can show that $\lambda_{iji}^{F\&F} = \lambda_{iji}^{D\&K}$: according to F&F, the kindness of Chap's reciprocation is measured with respect to the degree by which Chap alters Chip's actual from his expected payoff. Within each subgame of our mini-ultimatum game, Chap expects Chip to propose him an offer for which Chip seeks Chap's acceptance.[5] In turn, following D&K's arguments, acceptance is the only efficient strategy for Chap, and consequently acceptance leads to Chip's equitable payoff. Therefore, the difference between Chip's actual and equitable payoff measures within the context of our game whether Chap alters Chip's payoff by rejecting the offer.

[4] Again, we restrict ourselves to a slightly simplified version of D&K's model which – nonetheless – contains all crucial assumptions of this model.

[5] Notice that we can make this assumption since the Chip's anticipated immaterial utility component is zero: Chip's expected payoffs are Chap's alternations of Chip's payoffs which are anticipated by Chip. In other words, Chip cannot surprise himself in expectations. Therefore, if Chip offers Chap an offer for which Chip expects Chap's rejection, his expected utility from this offer is zero (no monetary and no immaterial utility). On the other hand, Chip would be better off by offering any alternative for which Chip assumes Chap's acceptance, since this yields the monetary utility. Thus Chip will never offer an alternative for which he anticipates a rejection.

The difference between the two approaches is settled in the specific form of $\kappa_{ij}^{F\&F}$. F&F separate the perceived (un)kindness into two terms, the outcome term Δ_j and the intention factor $\vartheta_j^{F\&F}$. The outcome term is formalized such that the evaluation of kindness is based on the inequity between Chip's and Chap's payoffs at a specific end node assuming that Chap chooses the efficient strategy:

$$\Delta_j = \pi_i - \pi_j \tag{6}$$

Again, within the context of our mini-ultimatum games, we can simply insert the payoffs following accepted offers into equation (6), since Chap believes Chip to expect Chap to accept the offer.

To derive perceived kindness, the outcome term is multiplied with the intention term, where reciprocal concerns come into play. Here, F&F distinguish between five payoff constellations from which different intentions are derived. More specifically, the intention factor accounts for Chip's intentional and unintentional choices depending on how Chip could have altered payoff constellation with regard to his own in combination with Chap's benefit:

$$\vartheta_j^{F\&F} = \begin{cases} 1 & \text{if } \pi_i^0 \geq \pi_j^0 \text{ and } \exists \tilde{\pi}_i \in \tilde{\Pi}_i : \tilde{\pi}_i < \pi_i^0, \\ \epsilon_i & \text{if } \pi_i^0 \geq \pi_j^0 \text{ and } \forall \tilde{\pi}_i \in \tilde{\Pi}_i : \tilde{\pi}_i \geq \pi_i^0, \\ 1 & \text{if } \pi_i^0 < \pi_j^0 \text{ and } \exists \tilde{\pi}_i \in \tilde{\Pi}_i : \tilde{\pi}_i > \pi_i^0 \text{ and } \tilde{\pi}_i \leq \tilde{\pi}_j \\ \max(1 - \frac{\tilde{\pi}_i - \tilde{\pi}_j}{\pi_j^0 - \pi_i^0}, \epsilon_i) & \text{if } \pi_i^0 < \pi_j^0 \text{ and } \exists \tilde{\pi}_i \in \tilde{\Pi}_i : \tilde{\pi}_i > \pi_i^0 \text{ and } \tilde{\pi}_i > \tilde{\pi}_j \\ \epsilon_i & \text{if } \pi_i^0 < \pi_j^0 \text{ and } \forall \tilde{\pi}_i \in \tilde{\Pi}_i : \tilde{\pi}_i \leq \pi_i^0, \end{cases} \tag{7}$$

where π_i^0, π_j^0 are payoffs resulting from accepting the specific offer in our design. Let $\tilde{\Pi}_i$ be the set of payoffs resulting from the acceptance of an alternative offer (but not the specific offer), $\tilde{\pi}_i$ be one element in $\tilde{\Pi}_i$, and ϵ_i be an individual parameter with $0 \leq \epsilon_i \leq 1$. This parameter is denoted as the pure outcome concern parameter. That is, ϵ_i measures Chap's unease with the inequity between Chip's and Chap's payoff, although Chip has no option to avoid the kind or mean offer.

The first two cases of (7) refer to intentions of Chip's actions favoring Chap moneywise: in the first one, Chip offers Chap not the smallest payoff possible although it is higher than his own one. This case is considered as fully intentional. In the second case, Chip offers Chap a higher payoff than his own one, but has no chance to avoid this. In this case, the outcome is nice, but Chip does not act intentionally so that the intention factor is reduced. The last three cases refer to negative intentions: the final case mirrors the second case into the negative domain; Chip offers Chap a smaller payoff than his own one, but has no chance to

avoid this. In this case, the outcome is mean, but Chip does not act intentionally so that the intention factor is reduced, whereas the third and fourth case refer to intentionally mean choices. In the fourth one, Chip offers Chap a smaller payoff than a possible alternative, but the offer is "somehow understandable" in the sense that the alternative yields less for Chip than for Chap. Therefore, Chip's intention is discounted according to the Chip's relative disadvantage under the alternative. In contrast, in the third case, Chip's unkindness is fully intentional, since there is a better alternative for Chap, which does not yield a lower payoff for Chip than for Chap.

Finally, $\kappa_{ij}^{F\&F}$ consists of the outcome term multiplied with the intention factor:

$$\kappa_{ij}^{F\&F} = \vartheta_j^{F\&F} \Delta_j \tag{8}$$

Put differently, to derive kindness in the intention factor model, intentions are "charged" by the difference between Chip's and Chap's monetary payoff, namely the outcome term. Inequity in favor of Chap increases the severity of Chip's kindness towards Chap, while disadvantageous inequity for Chap increases the severity of Chip's unkindness towards Chap. This means that the likelihood for Chap to behave unkindly (kindly) increases the more his normally resulting payoff falls below (surpasses) Chip's payoff, given that Chips chooses fully intentionally.

While F&F's (2006) model elegantly combines inequality aversion with reciprocal motivations, the formalization of the intention factor hosts strong, testable assumptions. Specifically, F&F's explicit quantification of the outcome term allows us to test the predictive success of their model:

Hyp$_{F\&F}$: *Decreasing the intention factor while keeping the outcome term constant across actions implies non-increasing rejection rates for those actions, while keeping the intention factor constant implies a constant rejection rate.*

3 The games

3.1 Design

To test both reciprocity models we design a series of twelve systematically varied mini-ultimatum games Γ_1 to Γ_{12} similar to the design by Falk et al. (2003). The proposer (Chip) receives throughout all games an endowment of at most 10 Taler.[6] In games Γ_1 to Γ_7, the proposer decides among two alternatives (GREEN,

[6]Notice that the total sum of Taler for alternative RED in Γ_5 and Γ_7 yields less than ten. Therefore, one may argue that as a consequence efficiency concerns could change the decisions in a systematic way in those games. As we will show in the result section, there are no indications

RED), in Γ_8 to Γ_{12} he decides among four alternatives on how to split the 10 Taler between the responder and herself (GREEN, RED, YELLOW, BLUE). The responder can either accept or reject the proposer's offer; in the former case, both parties reap their designated payoff, in the latter case both players receive zero Taler. In all twelve games one allocation is held constant at $(8, 2)$, while the remaining allocations differ depending on the purpose of each game. Table 1 lists all payoff allocations for the whole series of games.

	GREEN	RED	YELLOW	BLUE	$\kappa_{ij}^{D\&K}(8,2)$	$\vartheta_j^{F\&F}(8,2)$
Γ_1	8,2	8,2			0	ϵ_i
Γ_2	8,2	5,5			-1.5	1
Γ_3	8,2	9,1			0.5	ϵ_i
Γ_4	8,2	7,3			-0.5	1
Γ_5	8,2	4,3			-0.5	1
Γ_6	8,2	3,7			-2.5	$\max(\frac{2}{6}, \epsilon_i)$
Γ_7	8,2	3,4			-1.0	$\max(\frac{5}{6}, \epsilon_i)$
Γ_8	10,0	9,1	8,2	5,5	-0.5	1
Γ_9	9,1	7,3	8,2	5,5	-1.0	1
Γ_{10}	7,3	6,4	8,2	5,5	-1.5	1
Γ_{11}	9,1	10,0	8,2	6,4	0	1
Γ_{12}	9,1	7,3	8,2	6,4	-0.5	1

Table 1: *Payoff alternatives in* Γ_1 *to* Γ_{12} *along the values of* $\kappa_{ij}^{D\&K}$ *and* $\vartheta_j^{F\&F}$ *for the offer* $8, 2$ *according to the corresponding theories.*

Of course, non-reciprocal social preferences (e.g., inequity aversion) do not predict any difference concerning the likelihood to reject $8, 2$ across games (e.g., Fehr & Schmidt, 1999, and Bolton & Ockenfels, 2000, but also Levine, 1998; see Appendix A for further discussions). This changes substantially once we consider reciprocity. For our experiment, Γ_1 and Γ_2 serve as our baseline games. They provide benchmarks for our analysis in the sense that Γ_1 gives some indications for the responders' pure outcome concern (i.e., ϵ_i). That is, there is no intention involved in the offer GREEN in Γ_1, since the proposer has no alternative given that GREEN equals RED. In other words, rejecting in Γ_1 shows that the responder is ready to forgo 2 Taler, because he is so inequity averse that he does not want the proposer to gain 6 Taler more than him. Thus, rejections in Γ_1 indicate strong outcome concerns. On the other hand, Γ_2 shows us the response to a fully intentional, very unkind offer according to the two reciprocity models. That is, Γ_2 introduces a large distance between 2 and the equitable payoff ($\pi_j^{e_i} = 3.5$ in

that efficiency concerns influence decisions in Γ_5 and Γ_7, nor are the results of these games particularly important for the general results of our study.

Γ_2), while the offer is made with full intentions according to Falk & Fischbacher (2006, F&F).

With Γ_3 to Γ_7 we address both models in the context of simple games (i.e., games with two alternatives). According to F&F, Γ_3 depicts the fifth case of the intention factor resulting in an intention factor equal to the individual outcome concern parameter ϵ_i. As the outcome term is the same as in Γ_1, with reference to $\text{Hyp}_{F\&F}$ the rejection rate of GREEN is predicted to be equal to the rejection rate of Γ_1. Particularly, responders rejecting $8, 2$ in Γ_1 are expected to reject $8, 2$ in Γ_3 and vice versa.

Γ_4 and Γ_5 both depict a fully intentional decision context, corresponding to the third case of the intention factor. Thus, with reference to $\text{Hyp}_{F\&F}$ we predict equal rejection rates of GREEN for both games such that responders reject or accept $8, 2$ in both Γ_4 and Γ_5. Γ_6 and Γ_7 reverse payoffs of the two previous games in the alternative RED, resulting in a decision context characterized in the fourth case of the intention factor. Due to this structure, F&F's model predicts slightly smaller rejection rates of GREEN in Γ_7 (i.e., $\vartheta_j^{F\&F}$ is at least $\frac{5}{6}$ in this game), whereas the rejection rate in Γ_6 is predicted to be lower than in Γ_4, Γ_5 and Γ_7 – at least for those subjects whose choice indicated low outcome concerns (i.e., ϵ_i) by accepting GREEN and RED in Γ_1. Based on $\text{Hyp}_{F\&F}$, this implies on an individual basis that responders rejecting $8, 2$ in Γ_6 are expected to reject $8, 2$ in Γ_4, Γ_5 and Γ_7, while responders rejecting $8, 2$ in Γ_7 are expected to reject $8, 2$ in Γ_4 and Γ_5.

Let us now turn to the alternative theory: according to Dufwenberg & Kirchsteiger (2004, D&K), Γ_3 has a positive distance to the equitable payoff implying that rejections decrease utility. Therefore, responders do not reject $8, 2$ in Γ_3. Γ_4 and Γ_5 both depict the same distance to the equitable payoff suggesting equal rejection rates of GREEN for both games. In line with the predictions for F&F, responders rejecting $8, 2$ in Γ_4 are expected to reject $8, 2$ in Γ_5 and vice versa. However, D&K's predictions for Γ_6 and Γ_7 change substantially in comparison to F&F. Γ_6 introduces the most extreme distance to the equitable payoff within our sample of games, whereas Γ_7 introduces a smaller distance to the equitable payoff (though larger than in Γ_4 and Γ_5). It follows that Γ_6 has the largest rejection rate for $8, 2$, followed by Γ_2, Γ_7, and Γ_4 and Γ_5. This implies on an individual basis that responders rejecting $8, 2$ in Γ_4 and Γ_5 are expected to reject $8, 2$ in Γ_2, Γ_6 and Γ_7, while responders rejecting $8, 2$ in Γ_7 (Γ_2) are expected to reject $8, 2$ in Γ_2 and Γ_6 (Γ_6).

Γ_8 through Γ_{12} are designed to test the models in a richer context (i.e., games with more than two alternatives). Notice that the predictions according to F&F are constant as they depict the third case of the intention factor resulting in an intention factor equal to 1 for all games Γ_8 to Γ_{12}. In all games, Chip's choice of YELLOW is fully intentional due to the pairwise comparison between $8, 2$ and

$5,5$ (Γ_8 to Γ_{10}), or between $8,2$ and $6,4$ (Γ_{11} and Γ_{12}). In other words, the other alternatives except $5,5$ $(6,4)$ are irrelevant for determining the intention of choosing $8,2$. Hence according to $\text{Hyp}_{F\&F}$ the rejection rate for $8,2$ is equal across all five games, such responders either have to reject or accept $8,2$ in all games Γ_2, Γ_4, Γ_5 and Γ_8 to Γ_{12}.

Following D&K, Γ_8 through Γ_{10} represent a sequence of rising reference values implying increasing rejection rates for YELLOW according to $\text{Hyp}_{D\&K}$: reference values increase from 2.5 in Γ_8 to 3 in Γ_9 to 3.5 in Γ_{10} suggesting that rejection rates are expected to increase from Γ_8 through Γ_{10}. In turn, responders rejecting $8,2$ in Γ_8 are expected to reject $8,2$ in Γ_9 and Γ_{10}, while responders rejecting $8,2$ in Γ_9 are expected to reject $8,2$ in Γ_{10}.

Γ_{11} and Γ_{12} substitute the option BLUE allowing us some interesting comparisons within the richer games and across all games. According to $\text{Hyp}_{D\&K}$, the rejection rate of Γ_{12} is predicted to be equal to that of Γ_4, Γ_5 and Γ_8 (likewise, the rejection rate of Γ_7 is predicted to be equal to that of Γ_9, the rate for Γ_2 to be equal the rate of Γ_{10}). Responders rejecting $8,2$ in Γ_4, Γ_5, Γ_8 and Γ_{12} are expected to reject $8,2$ in Γ_7, Γ_9, Γ_2, Γ_{10} and Γ_6, as the distance to the equitable payoff is smaller in the former than in the latter games according to $\text{Hyp}_{D\&K}$. Likewise, substituting $5,5$ in Γ_8 against $6,4$ in Γ_{11} changes the character of $8,2$ according to D&K: $8,2$ is neither kind nor unkind in latter game implying no rejections of $8,2$ in Γ_{11}, but some rejections of this offer in Γ_8.

Summarizing our predictions, we rank (R) our games from those with the least likely rejection to the most likely rejection of $8,2$ according to D&K and F&F in Table 2 (i.e., games with a lower R are predicted to have less rejections than games with a higher R). Of course, this means on the within-subject level that responders who reject a game with a low R are predicted to reject all games with higher Rs as well. Notice that a rank of 0 results from D&K's prediction that no responder should reject $8,2$ in this game. Finally, the ranks of Γ_1, Γ_3, Γ_6 and Γ_7 depend on the individual parameter ϵ_i. As $1 \geq \epsilon_i \geq 0$, $\epsilon_i \approx 0$ for some players implies the ranks of 0, 0, 2, and 3, while $\epsilon_i \approx 1$ implies the ranks of 4, 4, 4, 4, respectively. As we are facing in the experiment a random sample, it seems plausible to assume in the aggregate the ranks of 1, 1, 2, and 3, whereas this need not be the case at the individual level.

3.2 Setting

The experiment is conducted as an online survey. Each participant plays every game in the role of either the responder, or the proposer. Subjects are randomly assigned to one of the two roles that they keep for the whole experiment. This allows us a within-subject analysis across games. On average, nine out of ten subjects participate as a responder, while approximately every tenth subject is

	GREEN	RED	YELLOW	BLUE	$R^{D\&K}(8,2)$	$R^{F\&F}(8,2)$
Γ_1	8,2	8,2			0	1
Γ_2	8,2	5,5			3	4
Γ_3	8,2	9,1			0	1
Γ_4	8,2	7,3			1	4
Γ_5	8,2	4,3			1	4
Γ_6	8,2	3,7			4	2
Γ_7	8,2	3,4			2	3
Γ_8	10,0	9,1	8,2	5,5	1	4
Γ_9	9,1	7,3	8,2	5,5	2	4
Γ_{10}	7,3	6,4	8,2	5,5	3	4
Γ_{11}	9,1	10,0	8,2	6,4	0	4
Γ_{12}	9,1	7,3	8,2	6,4	1	4

Table 2: *Payoff alternatives in* Γ_1 *to* Γ_{12} *along the game's rank according to the likelihood (least to most) of a rejection for* $8, 2$ *according to the corresponding theories.*

assigned to be the proposer. For responders, we apply the strategy method, in which responders have to accept or reject every possible payoff allocation (Selten, 1967). Consequently, each responder has to make 34 choices, each proposer 12 choices.

The experiment starts such that participants receive an invitation email including a link to access the online interface of the experiment. While accessing the interface, subjects are first familiarized with the procedure of the experiment and the instructions of the game on several pages on screen (Appendix B reports the instructions for the experiment). Participants are informed about all parameters of the game (including the payoff procedure) at this stage. Subsequently, participants submit all their choices without any feedback. The games are presented sequentially without the possibility to review earlier choices. The order of the games is randomized for each participant in order to exclude order effects. Payoffs in the experiment are denominated in Taler, that we exchange at 1 Taler for 2 Euros at the end of the experiment. During the experiment, all participants have on all decision screens the option to open an extra window showing again the instructions of the game. Finally, all participants have to fill out a short socio-demographic questionnaire. After the survey is completed by all participants, payment is determined from one randomly drawn game for every tenth formed pair of players. We randomly form pairs of one proposer and one responder each by pairing each proposer with one randomly drawn responder. The payoffs are computed according to the responder's decision corresponding to the particular choice of the proposer. Subjects are informed via email about their

payoff and pick up their earnings at the office of the experimental laboratory of the University of Hamburg.

In total, 496 students (various fields) from the University of Hamburg participated in two waves between November 2013 and March 2014 (each wave ran several days). 52.6 % percent were female, the median age was 25 years. We used hroot for recruitment (Bock et al., 2014). The average length of the entire online survey was approximately 20 minutes including instruction time, average payoff among the players receiving payoffs was 8.66 Euro implying an expected payoff for each player of 0.87 Euro.[7]

4 Results

4.1 Aggregate reciprocity

Let us start with proposers' decisions. We have 69 of them in our sample.[8] In general, the majority of proposers behaves very kindly such that the majority chooses the kindest offer in all games (the only exceptions are Γ_6 and Γ_7). In Γ_3 (and, trivially, in Γ_1), this is $8, 2$ (which is chosen by 86% of the proposers), while in the other simple games, $8, 2$ is offered by 25% (Γ_2), 16% (Γ_4), 41% (Γ_5), 61% (Γ_6) and 52% (Γ_7) of the proposers. In the richer games, the choice of $8, 2$ differs substantially over games: 22% of proposers offer $8, 2$ in Γ_8, 6% in Γ_9, 9% in Γ_{10}, 9% in Γ_{11} and 4% in Γ_{12}. Overall, it seems that $8, 2$ is not the most popular, but not an irrelevant alternative in all games.

Now, let us turn to responders' decisions. Table 3 reports the rejections rate of $8, 2$ in games Γ_1 to Γ_{12}.[9] As expected, non-reciprocal social preferences fail to characterize responders' decisions correctly. That is, neither are rejection rates similar across all games nor across the sub-sample Γ_1, Γ_2, Γ_4, Γ_5, Γ_6 Γ_7, and Γ_{10}. For instance, the rejection rates of Γ_2 and Γ_6, Γ_2 and Γ_{10}, and Γ_4 and Γ_7 are significantly different ($p < 0.002$).[10]

With regard to reciprocal preferences, there are some observation in line with both models for games Γ_1 to Γ_5. That is, out of 427 subjects, 235 reject $8, 2$ in Γ_4, and 231 subjects in Γ_5, so that – in line with both models – we cannot

[7]Expected (hourly) earnings correspond with previous experiments conducted via the internet or newspapers (e.g., Bosch-Domenech et al., 2002, Güth et al., 2003, 2007, Drehmann et al., 2005).

[8]Of course, their decisions are difficult to interpret as they are influenced by proposers' fairness considerations, but also anticipated fairness needs of responders. Nonetheless, we report our data in order to provide a complete picture of our experiment.

[9]The empirical rejection rates for alternatives GREEN and RED are identical in Γ_1.

[10]Throughout this subsection, we use a two-sided, paired t-test for the assessment of statistical significance.

reject the hypothesis that there are different rejection rates for GREEN in Γ_4 and Γ_5 ($p = 0.61$). However, contradicting $\text{Hyp}_{D\&K}$, 137 responders reject $8, 2$ in Γ_3. Furthermore, the rejection rate for Γ_2 (245 responders) is insignificantly different from the rate in Γ_4 ($p = 0.17$), and only weakly significantly different from the rate in Γ_5 ($p = 0.08$). Similarly, the rejection rates for Γ_6 (204 responders) and Γ_7 (209 responders) do not increase, but decrease significantly in comparison to Γ_4 ($p \leq 0.002$) and to Γ_5 ($p \leq 0.003$).

The results in Γ_4 and Γ_5 are in line with $\text{Hyp}_{F\&F}$. Yet, Falk & Fischbacher (2006, F&F) fail to describe rejections in games with limited intention factors. That is, concerning the rejection rates in Γ_6 and Γ_7, we expect a smaller number in the first than in the second game (although there are some limitations to this expectations if ϵ_i is close to one for the majority of players; see our earlier comment at the end of Section 3.1). In contrast, there is no significant difference between the rates for both games ($p = 0.59$). However, this result holds, even if we restrict our focus on those responders who did not reject in Γ_1 (i.e., subjects without pronounced outcome concerns). Out of 267 responders who accepted both offers in Γ_1, 71 (70) rejected $8, 2$ in Γ_6 (Γ_7); there is no significant difference between the two rejection rates ($p = 0.8$). Likewise, the rejection rates of Γ_1 (152 responders) and Γ_3 differ weakly significant ($p = 0.08$) despite the predictions of $\text{Hyp}_{F\&F}$.

Nonetheless, our overall results suggest that F&F characterizes behavior more accurately. That is, rejection rates in the simple games follow the predictions of F&F – particularly for fully intentional decisions, while they are poorly described by Dufwenberg & Kirchsteiger (2004, D&K).

Next, we look at the rejection rate of $8, 2$ in Γ_8 through Γ_{12}. We observe two "blocks" of games with respect to their rejection rates. Γ_9, Γ_{10} and Γ_{12} have all rejection rates of approximately 0.6, while Γ_8 and Γ_{11} have rejection rates of approximately 0.5. All rejection rates in the first block (260/263/251) are significantly higher than rejection rates in the second block (214/205; $p < 0.001$), while within blocks, there is only one weakly significant difference between Γ_{10} and Γ_{12} ($p = 0.08$, all other comparisons $p \geq 0.18$). Hence, contradicting $\text{Hyp}_{D\&K}$, there is little evidence that the sequence of rising distances to the reference value triggers rejections in a systematic way, whereas the same distance to the reference value leads to significantly different rejection rates between Γ_8 and Γ_{12} ($p < 0.001$). Likewise, a comparison across simple and richer games show significant different rejection rates between Γ_1 and Γ_{11} ($p < 0.001$).

Similarly, predictions according to $\text{Hyp}_{F\&F}$ are misaligned with results due to the lower rejections rate of $8, 2$ in Γ_8 and Γ_{11} compared to Γ_9, Γ_{10} and Γ_{12}. In contrast to our earlier results, which corroborate with F&F's model, we provide here evidence for a lack of generality (or a limiting specificity) of F&F's model. More precisely, by changing the structure of the games in a way that we add

	Rejection rates for $8,2$	$R^{D\&K}(8,2)$	$R^{F\&F}(8,2)$
Γ_1	0.35	0	1
Γ_2	0.57	3	4
Γ_3	0.32	0	1
Γ_4	0.55	1	4
Γ_5	0.54	1	4
Γ_6	0.48	4	2
Γ_7	0.49	2	3
Γ_8	0.50	1	4
Γ_9	0.61	2	4
Γ_{10}	0.62	3	4
Γ_{11}	0.48	0	4
Γ_{12}	0.59	1	4

Table 3: *Rejection rates for* $8,2$ *in* Γ_1 *to* Γ_{12} *along the game's rank according to the likelihood (least to most) of a rejection for* $8,2$ *according to the corresponding theories.*

a number of alternatives, reciprocal behavior cannot be satisfactorily described by F&F intention-based model anymore. Interestingly, the comparison across simple and richer games also casts some doubts onto F&F's model (e.g., Γ_4 and Γ_{12}: $p = 0.04$; or Γ_2 and Γ_8: $p < 0.001$), although in those cases the alternatives of the simple games are subsets of the alternatives in the richer games. We discuss this point in greater detail in the next section.

Given our observations, it seems that neither D&K's model nor F&F's model characterize rejection behavior in the richer games accurately. Particularly, both models fail to predict the occurrence of the two blocks. The results suggest that both models miss to incorporate an important characteristic of reciprocity.

4.2 Individual reciprocity

In the following, we test for the consistency of our results on an individual level. That is, we test the personal implications of both models across games. Let us start with D&K in the simple games: in accordance with $\text{Hyp}_{D\&K}$, we cannot reject the hypothesis that the same subjects reject Γ_4 and Γ_5 ($p = .61$).[11] At the same time, we have to reject the hypothesis that at least all responders rejecting $8,2$ in Γ_4 and Γ_5 (203 responders do so) reject $8,2$ in Γ_2, Γ_6 and Γ_7 as well ($p < 0.001$): only 151 responders reject $8,2$ in all five games. Likewise,

[11] Throughout this subsection, we use a two-sided Friedman test for the assessment of statistical significance.

the prediction that responders rejecting $8,2$ in Γ_7 reject $8,2$ in Γ_2 and Γ_6 is not supported by the data ($p < 0.001$): only 160 responders reject $8,2$ in all three games. Finally, the prediction that responders who reject $8,2$ in Γ_2 do so in Γ_6 as well is also not supported by the data ($p < 0.001$): 186 responders reject $8,2$ in both games.

Turning to F&F, we find that 107 responders reject $8,2$ in Γ_1 and Γ_3 (out of 152/137 rejecting Γ_1/Γ_3), so that there is weakly significant evidence that not the same subjects reject this offer in both games ($p = 0.08$). Similarly, $\text{Hyp}_{F\&F}$ is not supported in the sense that from 204 (209) responders rejecting $8,2$ in Γ_6 (Γ_7), 153 (180) reject the same offer in Γ_4, Γ_5 and Γ_7 (Γ_4 and Γ_5). Here, we have to reject the hypothesis that the same subjects reject this offer in all four games ($p < 0.001$, and $p = 0.002$, respectively). However, we cannot discard the hypothesis that the same subjects reject $8,2$ in Γ_4 and Γ_5 ($p = 0.61$): from 235 (231) rejecting $8,2$ in Γ_4 (Γ_5), 203 reject the offer in both games. In other words, F&F organizes the data well, if offers are fully intentional referring to their model.

For the richer games, D&K predict that responders rejecting $8,2$ in Γ_8 (Γ_9) are expected to reject the same offer in Γ_9 and Γ_{10} (Γ_{10}). There is little evidence for the claims: from 214 (260) responders rejecting $8,2$ in Γ_8 (Γ_9), 194 (236) responders do so in Γ_9 and Γ_{10} (Γ_{10}) as well ($p < 0.001$ for both comparisons). Likewise, not the same responders reject $8,2$ in Γ_4, Γ_5, Γ_8 and Γ_{12} ($p < 0.001$), Γ_7 and Γ_9 ($p < 0.001$), and Γ_2 and Γ_{10} ($p = 0.01$). Finally, from 167 responders rejecting $8,2$ in Γ_4, Γ_5, Γ_8 and Γ_{12}, 137 reject $8,2$ in Γ_7, Γ_9, Γ_2, Γ_{10} and Γ_6 as well. Again, the claim that the same responders reject $8,2$ across all those games is not supported ($p < 0.001$).

We conclude our result section by testing F&F in the context of richer games: according to $\text{Hyp}_{F\&F}$, the same responders reject $8,2$ in games Γ_2, Γ_4, Γ_5 and Γ_8 to Γ_{12}. This claim is not supported by the data ($p < 0.001$). Even if we restrict our analysis to games Γ_8 to Γ_{12}, there is little evidence that the same responders reject $8,2$ in those games ($p < 0.001$). However, we cannot reject the claim that the same responders reject $8,2$ in Γ_9, Γ_{10} and Γ_{12} ($p = 0.19$). Thus there seems to be an important difference between Γ_9, Γ_{10} and Γ_{12} on the one hand, and Γ_8 and Γ_{11} on the other that influences reciprocity substantially in our experiment.

5 Discussion

Our results indicate for both models weaknesses when predicting perceived kindness. Dufwenberg & Kirchsteiger (2004, D&K) provide with their $\kappa_{ij}^{D\&K}$ a partition in the degree of kindness which is too detailed and cannot predict behavior. Falk & Fischbacher's (2006, F&F) kindness measurement yields a partition of kindness which is too general in richer settings due to the pairwise comparison

of alternatives. Therefore, we want to discuss a potential combination of both reciprocity models: On the one hand, we want to simplify F&F's model so that it is applicable in richer environments while keeping at least to a large extend its predictive power in simple games. On the other hand, we want to incorporate the global assessment of alternatives' kindness to provide sufficient partition of kindness, and, consequently, better predictive power in richer settings.

We would like to stress that we do not attempt to present a fully elaborated model, but we want to sketch a possible avenue for the further development of research on reciprocity. Moreover, we have to admit that our modification does not handle the intra-personal inconsistences which have been demonstrated in the last subsection. Again, we would consider our approach with respect to this work as a general starting point which needs further elaboration.

As our data corroborate in simple games F&F's distinction between five generic cases of intention (and the distinction between the intention and the extent to which this alternative is considered to be kind or unkind), we do not seek to modify this feature.[12] Hence, we propose a κ_{ij}^{new} which consists of the product of an intention factor and the outcome term Δ_j according to equation (6). However, the major extension is a global assessment of j's alternatives. Not only do our experimental results suggest this approach, but we consider it as unrealistic that in more complex situations – and we study in our experiment complexity only to the extent that we test games with four instead of two alternatives – the existence of one clearly (un)kind action determines the choice of another alternative as a fully intentional act of (un)kindness.

Thus, similar to D&K's approach, we model intention relative to some reference value. Specifically, the reference value we propose, the median offer (i.e., the "middle" payoff within the set of corresponding end notes resulting from the consecutive choice of efficient strategies) provides the additional benefit of being robust against outliers in the set of alternatives (D&K discuss the problem of outlying payoffs extensively in their paper). Formally, let us denote with π_i^M i's median payoff among the set of payoffs resulting from j's choice of efficient strategies.[13] Then, we define the intention factor of j's move ϑ_j^{new} of offering a

[12] Again, we would like to stress that we find no significant differences in terms of rejections between Γ_4 and Γ_5 as well as Γ_6 and Γ_7, but significant differences between both blocks of games. On a side-note, we consider the fact that there is no significant difference of rejection rates within both blocks – although one game per block yields in sum less than 10 Taler – as evidence that efficiency concerns are less important in this setting.

[13] We define π_i^M implicitly with respect to the cumulative distribution function $F(x)$ on i's set of payoffs resulting from i's and j's choice of any combination of efficient strategies in a game: π_i^M satisfies both inequalities $\int_{(-\infty,\pi_i^M]} dF(x) \geq 0.5$ and $\int_{[\pi_i^M,\infty)} dF(x) \geq 0.5$.

specific payoff combination π_i^0, π_j^0 as follows:

$$\vartheta_j^{new} = \begin{cases} 1 & \text{if } \pi_i^0 > \pi_i^M \text{ and } \pi_i^0 > \pi_j^0, \\ \epsilon_i & \text{if } \pi_i^0 \leq \pi_i^M \text{ and } \pi_i^0 \geq \pi_j^0, \\ \epsilon_i & \text{if } \pi_i^0 \geq \pi_i^M \text{ and } \pi_i^0 \leq \pi_j^0, \\ 1 & \text{if } \pi_i^0 < \pi_i^M, \pi_i^0 < \pi_j^0, \text{ and } \exists \tilde{\pi}_j \in \tilde{\Pi}_j : \tilde{\pi}_j > \pi_i^M \\ \max(1, 2\epsilon_i) & \text{if } \pi_i^0 < \pi_i^M, \pi_i^0 < \pi_j^0, \text{ and } \forall \tilde{\pi}_j \in \tilde{\Pi}_j : \tilde{\pi}_j \leq \pi_i^M, \end{cases} \tag{9}$$

where $\tilde{\Pi}_j$ be the set of payoffs resulting from the acceptance of an alternative offer (but not the specific offer), and $\tilde{\pi}_j$ be one element in $\tilde{\Pi}_j$.

That is, like F&F's approach, our intention factor differentiates between five categories of i's outcomes resulting from j's action: payoffs implying smaller payoffs for j than for i are considered as fully intentionally kind if they are larger than the reference value, whereas they are accidentally kind if they are smaller or equal to the reference value. In turn, there are accidentally unkind offers which imply larger payoffs for j than for i if they are larger or equal to the reference value. Finally, j's action leading to i's payoff being smaller than i's reference value and j's payoff is considered to be fully intentional unkind only if j could choose better alternatives for himself. That is, if the unkind offer is "somehow understandable" in the sense that all other alternatives yields less for j than i's reference value, the offer is still perceived as intentionally unkind but not that much. Only, if there is at least one other alternative which yields for j more than i's reference value, choosing the specific alternative is fully intentional (and unkind).

Notice that the latter two cases translate F&F's observation that "the perception of the unfair offer depends on how much j has to sacrifice in order to make the more friendly offer" (F&F, 2006, p. 297) into the context of a global assessment of i's payoffs. That is, if making a more friendly offer than $8, 2$ implies that j earns less than i's reference value, this offer is still unkind, but with limited intention.

Based on our reformulation of ϑ_j^{new}, we obtain a new rank order for the likelihood of a rejection for $8, 2$ which is reported in Table 4: the predictions based on ϑ_j^{new} follow qualitatively the one based on $\vartheta_j^{F\&F}$ for the simple games. However, using ϑ_j^{new} one can predict the two blocks of rejection rates in the richer games. Likewise, the rejection rates for $7, 3$ in Γ_9, Γ_{10} and Γ_{12}: in Γ_9 and Γ_{12} where $7, 3$ is unfavorable but mildly unkind according to $\kappa_{ij}^{new} = \vartheta_j^{new}\Delta_j$, 101 and 106 responders reject the offer, 151 do so in Γ_{10} where this offer is fully intentionally unkind.[14]

[14] $p < 0.001$ for the hypothesis that the rejection rate in Γ_{10} equals one in the other two games, whereas $p = 0.466$ for the hypothesis that the rejection rates in Γ_9 and Γ_{12} are the

Thus, κ_{ij}^{new} allows us to combine F&F's idea of a kindness term which differentiates between the intention of an action and the extend of kindness with D&K's approach to assess an entire game by means of a reference value. We would like to stress that this seems for us particularly important when making predictions in the context of richer decision environments. For instance, in an ultimatum-type game with several alternatives, the assumption that players evaluate the kindness of a specific offer by pairwise comparisons between alternatives seems unrealistic to us at least due to the shire computational effort it takes to compare alternatives against each other. Rather, we follow D&K's idea that players condense alternatives by means of reference values. As such, our model is located in some sense half way between the models by D&K and F&F: it processes more information than the model by D&K. On the other hand, our approach generalizes over alternatives by a larger extend than F&F by forming reference values. Whether our modifications optimize the tradeoff between the generalisability of the model to various situations and the accurate prediction of specific behavior is an open question and requires future research.

	Rejection rates for $8, 2$	$\vartheta^{new}(8, 2)$	$R^{new}(8, 2)$
Γ_1	0.35	ϵ_i	1
Γ_2	0.57	1	3
Γ_3	0.32	ϵ_i	1
Γ_4	0.55	1	3
Γ_5	0.54	1	3
Γ_6	0.48	$\max(1, 2\epsilon_i)$	2
Γ_7	0.49	$\max(1, 2\epsilon_i)$	2
Γ_8	0.50	ϵ_i	1
Γ_9	0.61	1	3
Γ_{10}	0.62	1	3
Γ_{11}	0.48	ϵ_i	1
Γ_{12}	0.59	1	3

Table 4: *Rejection rates for* $8, 2$ *in* Γ_1 *to* Γ_{12} *along the game's rank according to the likelihood (least to most) of a rejection for* $8, 2$ *according to the modification of* κ_{ij}.

6 Conclusion

Although reciprocity is one fundamental cornerstone of human behavior, modeling reciprocity still is a challenge for social scientists. The current study analyzes

same.

two of the most established approaches, the reference value model by Dufwenberg & Kirchsteiger (2004, D&K) and the intention factor model by Falk & Fischbacher (2006, F&F). We point out that there are two major differences between the two approaches: the first model measures perceived kindness of an action in relation to a reference value while the second model distinguishes between the intention of an action and the extent to which the action is perceived as being unkind or kind. The latter element of F&F's model relies on the inequity of the proposed payoffs whereas the former element results from a pairwise comparison between alternatives.

We test both models within the context of mini-ultimatum games with two and four alternatives. Results show that F&F's approach works fine in the games with two alternatives, but has important drawbacks in the games with four alternatives, both with respect to the average numbers but also once we run a within-subject analysis. On the other hand, D&K's model fails to characterize behavior within both contexts. From this we conclude that D&K idea to measure perceived kindness in one variable, the distance to the equitable payoff, does not sufficiently capture the nature of perceived (un)kindness. Likewise, the pairwise comparison seems to lose its predictive power once we depart from simple games with two alternatives. Therefore, we present and discuss a potential modification of F&F's reciprocity model which includes elements of D&K'S approach.

To conclude, more research is needed to model reciprocity in a sufficient way. Perhaps, the question is not whether there is a true model mapping reciprocity, but whether there is a model that adequately balances the need for generalisability across different games with a satisfactory good predictability of behavior within a specific environment. Elsewhere, it has been shown that reciprocity itself encompasses a number of different subtypes of social utility (e.g., Nicklisch & Wolff, 2012). Therefore, we have to ask ourselves whether we want to model the behavior in one specific game which may trigger one specific form of reciprocity, or whether we want to rely on a general model, which, however, has less predicting power in special situations. The answer to this question we cannot provide here. Therefore, we would like to invite future research to follow this avenue, or, maybe, prove it wrong.

References

[1] Ambrus, A. & B. Greiner (2012), Imperfect public monitoring with costly punishment: An experimental study, *American Economic Review* **102**, 3317–3332.

[2] Andreoni, J., M. Castillo & R. Petrie (2003), What do bargainers' preferences look like? Experiments with a convex ultimatum game, *American Economic Review* **93**, 672–685.

[3] Battigalli, P. & M. Dufwenberg (2007), Guilt in games, *American Economic Review* **97**, 170–176.

[4] Bock, O., A. Nicklisch & I. Baetge (2014), hroot: Hamburg registration and organization online tool, *European Economic Review* **71**, 117–120.

[5] Bolton, G.E. & R. Zwick (1995), Anonymity versus punishment in ultimatum bargaining, *Games and Economic Behavior* **10**, 95–121.

[6] Bolton, G.E. & A. Ockenfels (2000), ERC: A theory of equity, reciprocity, and competition, *American Economic Review* **90**, 166–193.

[7] Bosch-Domenech, A., J.G. Montalvo, R. Nagel & A. Satorra (2002), One, two,(three), infinity, ...: Newspaper and lab beauty-contest experiments, *American Economic Review* 92, 1687-1701.

[8] Cox J.C. & C. Deck (2003), On the nature of reciprocal motives, *Economic Inquiry* **41**, 20—26.

[9] Cox, J.C., D. Friedman & S. Gjerstad (2007), A tractable model of reciprocity and fairness, *Games and Economic Behavior* **59**, 17–45.

[10] Cox, J.C., D. Friedman & V. Sadiraj (2008), Revealed Altruism, *Econometrica* **76**, 31–69.

[11] Dhaene, G. & J. Bouckaert (2010), Sequential reciprocity in two-player, two-stage games: An experimental analysis, *Games and Economic Behavior* **70**, 289–303.

[12] Drehmann, M., J. Oechssler & A. Roider (2005), Herding and contrarian behavior in financial markets: An internet experiment, *American Economic Review* **95**, 1403-1426.

[13] Dufwenberg, M. & G. Kirchsteiger (2004), A theory of sequential reciprocity, *Games and Economic Behavior* **47**, 268–298.

[14] Falk, A. & U. Fischbacher (2006), A theory of reciprocity, *Games and Economic Behavior* **54**, 293–315.

[15] Falk, A., E. Fehr & U. Fischbacher (2003), On the nature of fair behaviour, *Economic Inquiry* **41**, 20–26.

[16] Falk, A., E. Fehr & U. Fischbacher (2008), Testing theories of fairness – Intentions matter, *Games and Economic Behavior* **62**, 287–303.

[17] Fehr, E. & K.M. Schmidt (1999), A theory of fairness, competition, and cooperation, *The Quarterly Journal of Economics* **114**, 817–868.

[18] Gächter, S. & C. Thöni (2007), Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis, *Journal of Economic Behavior & Organization* **76**, 531–543.

[19] Güth, W., C. Schmidt & M. Sutter (2003), Fairness in the mail and opportunism in the internet: A newspaper experiment on ultimatum bargaining, *German Economic Review* 4, 243-265.

[20] Güth, W., C. Schmidt & M. Sutter (2007), Bargaining outside the lab – a newspaper experiment of a three-person ultimatum game, *The Economic Journal* 117, 449-469.

[21] Levine, D. (1998), Modeling altruism and spitefulness in experiments, *Review of Economic Dynamics* **1**, 593–622.

[22] Nicklisch, A. & I. Wolff (2012), On the nature of reciprocity: Evidence from the ultimatum reciprocity measure, *Journal of Economic Behavior & Organization* **84**, 892–905.

[23] Pelligra, V. (2011), Reciprocating kindness: An experimental investigation, *Working Paper*.

[24] Rabin, M. (1993), Incorporating fairness into game theory and economics, *American Economic Review* **83**, 1281–1302.

[25] Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: H. Sauermann (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen: J.C.B. Mohr, 136–168.

[26] Sobel, J. (2005), Interdependent preferences and reciprocity, *The Journal of Economic Literature* **43**, 392–436.

[27] Stanca, L., L. Bruni & L. Corazzini (2009), Testing theories of reciprocity: Do motivations matter? *Journal of Economic Behavior & Organization* **71**, 233–245.

Appendix A: Predictions according to outcome concerned social utility

Inequity aversion

Given inequity averse preferences, we have to claim that responders either accept or reject $8, 2$ in all games Γ_1 to Γ_{12}. The reason for this is rather obvious: as the same offer $8, 2$ is considered throughout all games and inequity aversion preferences take only the outcome of a proposal into consideration, there is no difference in the utility resulting from acceptance across games, nor from rejection across games.

Levine's (1998) model

In the following we want to derive a prediction for behavior in our games according to Levine's (1998) model. We choose this model, since it can be characterized as some intermediate step between models based on outcome concerns and reciprocity models. The reason for this is that the proposer partly reveals his taste for altruism through his choice among the alternatives of the game. This information updates the weight for altruism in the responder's utility function and may lead to rejections if altruism is negative. Formally, we can define the utility function of proposer i (paired with responder j) according to Levine (1998) as:

$$U_i^L = \pi_i(a_i, a_j) + \frac{\alpha_i + \varrho_i \alpha_j}{1 + \varrho_i} \pi_j(a_i, a_j)$$

where $\pi_i(a_i, a_j)$ is i's monetary payoff of an offer, while α_i is i's taste for altruism ($-1 < \alpha_i < 1$); finally, ϱ_i measures the importance of j's altruism for i's utility ($0 \leq \varrho_i \leq 1$). Of course, α_i and ϱ_i are i's private information. However, by choosing a specific offer in the game, the proposer i partly reveals his task for altruism to the responder j who updates his utility U_j^L accordingly. Therefore, j utility changes with i's choice of alternatives in the game. Yet, it turns out that i's choice of $8, 2$ is uninformative in Γ_1, Γ_2, Γ_4, Γ_5, Γ_6, Γ_7, and Γ_{10} in the sense that it only reveals $a_i < 1$ which is known from the beginning. As an example, consider Γ_2. Here, choosing $8, 2$ implies $8 + \frac{\alpha_i + \varrho_i \alpha_j}{1+\varrho_i} 2 > 5 + \frac{\alpha_i + \varrho_i \alpha_j}{1+\varrho_i} 5$ which is equivalent to $1 + \varrho_i(1 - \alpha_j) > \alpha_i$. It follows that the maximum of α_i is 1.

Thus the choice of a specific offer in our mini-ultimatum games does not result in an update of j's belief concerning a_i. It follows that in all of the previously mentioned games j rejects the offer of $8, 2$ if $0 > 2 + \frac{\alpha_j + \varrho_j \alpha_i}{1+\varrho_j} 8$. It follows that j rejects if $\alpha_j < \hat{\alpha}$ with $\hat{\alpha} \in (-1, \ldots, 0.5]$ depending on j's specific ϱ_j. That is to say, if a responder rejects $8, 2$ in one of the games Γ_1, Γ_2, Γ_4, Γ_5, Γ_6, Γ_7, or Γ_{10},

he should reject $8, 2$ in all seven games. Following the same rational, Levine's model predicts that i does not offer $8, 2$ in Γ_3, Γ_8, Γ_9, Γ_{11}, and Γ_{12}. Therefore, we are hardly able to form predictions with respect to responser's behavior, and do not discuss the results of these games in the context of Levine's model.

Appendix B: Experimental instructions

Welcome to the experiment!

In the following you will participate in a game in which you can earn a considerable amount of money depending on your decisions and the decisions of other participants. Therefore we kindly ask you to read the following instructions and after that make your decisions. Completing the experiment takes about 20 minutes.

In the experiment we talk of *Taler*. At the end of the experiment we will exchange the *Taler* to Euro, with 1 *Taler* = 2 Euro. We will inform you after the experiment via e-mail, if you will receive a payoff.

On the next page we will explain the rules of the experiment.

This experiment consists of twelve games, in which two participants interact with each other.

We call both persons Player A and Player B. At the beginning of the experiment you will randomly be assigned the role of one player and keep that role during the whole experiment. In every game Player A decides on the allocation of an endowment of 10 Taler at a maximum. There are either two or four alternatives of how to divide the endowment. By deciding for one alternative, Player A offers this alternative to Player B. At the same time, Player B decides for each alternative of a game, if she accepts or declines this offer. If Player B declines, both players earn 0 Taler. After you have made your decisions for every game, you will randomly be assigned a player of the opposing role.

Your earnings will be determined by your decision and the decision of your opponent for one randomly drawn game. This means, that each of your decisions is equally relevant for your earning. We pay out every tenth pair of a Player A and Player B.