HOLD-UP: WITH A VENGEANCE

MARTIN DUFWENBERG, ALEC SMITH and MATT VAN ESSEN*

When contracts are incomplete or unenforceable, inefficient levels of investment may occur because of hold-up. If individuals care for negative reciprocity, these problems may be reduced, as revenge becomes a credible threat. However, negative reciprocity has this effect only when the investor holds the rights of control of the investment proceeds. We explore this issue analytically, deriving predictions for hold-up games which differ as regards assignment of rights of control. We also test and support these predictions in an experiment. (*JEL* C72, C92, D23, L14)

I. INTRODUCTION

The hold-up literature shows how relationship-specific investments and incomplete contracts combine to hurt partnership profitability.[1] The conclusions typically build on the assumption that agents selfishly maximize own income. However, hold-up scenarios involve opportunistic exploitation. Intuition as well as a wealth of evidence suggest exploited parties may get irritated and strike back. It may seem such negative reciprocity can deter exploitation and render hold-up less problematic.

*We thank Gary Charness, Yan Chen, Jim Cox, Andrew Daughety, Tore Ellingsen, Ben Hermalin, Georg Kirchsteiger, Jennifer Reinganum, Uzi Segal, Joel Sobel, an anonymous referee, and participants of several seminars for helpful comments, and Bruce Willis and his 1995 team for inspiration for the title. Dufwenberg's research was supported by the NSF (grant # SES-0921929).

Dufwenberg: Department of Economics, University of Arizona, Tucson, AZ 85721, and Department of Economics, University of Gothenburg, Sweden. Phone 520-626-1540, Fax 520-621-8450, E-mail martind@eller.arizona.edu

Smith: Division of Humanities and Social Sciences, California Institute of Technology, Pasadena CA 91125. Phone 626-395-5967, Fax 626-405-9841, E-mail acs@hss.caltech.edu

Van Essen: Department of Economics, Finance, and Legal Studies, University of Alabama, Tuscaloosa, AL. Phone 719-510-5502, Fax 205-348-0590, E-mail mjvanessen@cba.ua.edu

1. The idea can be traced back at least to Williamson (1975, 1985) and Klein, Crawford, and Alchian (1978). See Grout (1984) and Tirole (1986) for early formal theory, Tirole (1988, pp. 24–27, especially Example 1) and Hart (1995, pp. 73–85) for textbook style introductions, and Che and Sakovics (2008) for a recent review covering much modern developments. Hold-up analysis is an Archimedean point for contract theorists who, starting with Grossman and Hart (1986) and Hart and Moore (1990), explain how organizational forms and contractual arrangements arise to reduce the underinvestment hold-up may cause.

We explore the issue analytically and experimentally. We show that whether negative reciprocity mitigates hold-up depends predictably on the nature of the investment. Two contrasting examples (inspired by question 10 in chapter 5 of Besanko et al. 2010) illustrate the key principle involved:

EXAMPLE 1. An artist (player 1) has been asked by a presumptive buyer (player 2) to paint a "beautiful portrait of 2." 1 may disagree or agree. In the former case, 1 and 2 go separate ways. In the latter case, 1 spends $2,000 worth of his/her time on the painting, and a contract says 2 should subsequently pay $5,000 to 1. The value to 2 is $8,000, but 2 may complain and claim (falsely) that the portrait is "rather ugly" and attempt to renegotiate offering a new price of $1,000. Given the ambiguity of what constitutes beauty, 1 cannot enforce the $5,000 payment and will have to accept or reject the new offer. 1 knows that no person other than 2 would pay to acquire the painting.

Here, 2 could shoot himself in the foot by trying to renegotiate the price. 1 might become angry or, to use a phrase consistent with the intentions-based reciprocity theories of Rabin (1993) or Dufwenberg and Kirchsteiger (2004) (D&K), deem 2 unkind and desire to be unkind in return. 1 would then prefer to reject the renegotiated offer, and destroy (or disfigure and

ABBREVIATIONS

SRE: Sequential Reciprocity Equilibrium
VE: Vengeance Equilibrium

Economic Inquiry
(ISSN 0095-2583)

doi:10.1111/j.1465-7295.2010.00366.x

exhibit) the portrait. If 2 foresees this, then we have a case where negative reciprocity mitigates hold-up.

Our second example suggests that this insight only carries so far:

EXAMPLE 2. A professor (player 1) has been asked by a student (player 2) to extracurricularly teach 2 "how to use game theory to make a lot of money." 1 may disagree or agree. In the former case, 1 and 2 go separate ways. In the latter case, 1 spends $2,000 worth of his time talking to 2, and a contract says 2 should subsequently pay $5,000 to 1. The value to 2 is $8,000, but 2 may instead complain and claim (falsely) that the tutoring "was only good enough that he/she can make a moderate amount of money" and offer a renegotiated price of $1,000. Given the ambiguity of what is "a lot," 2 cannot enforce the $5,000 payment and he will have to accept or reject the new offer.

As strategic structure and monetary payoffs go, Example 2 may seem similar to Example 1. The sole difference is that if player 2 proposes to renegotiate the price, and if 1 rejects the offer, then 2 gains rather than loses. This reflects how education, unlike a portrait, comprises human capital which cannot be withheld. When players are vengeful this difference has repercussions for the entire strategic analysis. There is no way for player 1 to hurt player 2, therefore less to deter 2 from proposing a renegotiated price, and therefore less incentive for 1 to agree to the tutoring. Even if players are motivated by negative reciprocity, hold-up remains an issue.[2]

The key difference between Examples 1 and 2, however, is not whether a relation-specific investment concerns human capital. Rather, the issue concerns (to use a term of Grossman and Hart's 1986) who has "residual rights of control" of the proceeds of the investment. Human capital may be a prominent source of residual rights of control, but other sources are possible too. Say, to make things concrete, that we had an example (suggested by Ben Hermalin) with a home-owner and a plumber who installs a new system of pipes and drains in the home-owner's bathroom. Whether this situation most resembles Example 1 or Example 2 depends on whether legislation makes it feasible for the plumber to pull out and destroy the pipes and drains if the home-owner tries to renegotiate a preagreed price.

Examples 1 and 2 were selected for pedagogical clarity, not because they were the economically most significant we could come up with. However, these cases are structurally similar to grander scenarios, with or without transfer from seller to buyer of residual rights of control. This could involve tailor-made multimillion dollar equipment for space-walks (where the seller keeps control) or expensive training-programs for key personnel in the oil industry (involving transfer of human capital), etc. In some cases, the held-up party may be a firm and it is reasonable to wonder whether firms are as prone to reciprocation as individuals. We suggest so, because at the end of the day firms are run by individuals. Hart (2008) makes the same point, suggesting that "[l]arge corporations are run by individuals who have big egos and presumably therefore can have strong emotions."

This paper explores whether the intuitions outlined in connection with Examples 1 and 2 prove logically valid when a formal model (a D&K-modification) of negative reciprocity is applied. They are, up to a caveat concerning twin equilibria. Theoretical stories, however, intriguing per se, gain in value if they possess empirical relevance and in order to examine whether we can thus boost confidence in our predictions, we also designed an experiment to test them. We report the results, which are supportive.

Several contributions (spanning theory and experiments) have suggested that fairness or reciprocity may be relevant to hold-up although they explore other topics than highlighting connections between residual rights-of-control and vengeance using intentions-based reciprocity theory.[3] Most closely related are Ellingsen and Johannesson (2004a) and von Siemens (2009), who study hold-up games reminiscent of our Example 1 and examine, respectively, preplay communication and forms of incomplete information. Our approach differs in style from some preceding studies (including Ellingsen

2. Our focus here, and in the remainder of the paper, concerns *negative* reciprocity, where unkindness breeds unkindness in return. Hold-up problems may also be affected by *positive* reciprocity, where kindness breeds kindness. In Section III, we discuss this further and make a case for not focusing on positive reciprocity.

3. See Hackett (1994), Gantner, Guth, and Konigstein (2001), Ellingsen and Johannesson (2004a, 2004b), Sobel (2005, section 4.2.1), Sloof, Sonnemans, and Oosterbeek (2007), Hart and Moore (2008), Fehr, Hart, and Zehnder (2008), Hart (2008), Fehr, Kremhelmer, and Schmidt (2008), and von Siemens (2009).

and Johannesson 2004a; Fehr, Kremhelmer, and Schmidt 2008) which start with some clever experimental design incorporating hold-up and then offer "behavioral explanations" or "a theoretical interpretation" (respectively: p. 407, p. 1263). We reverse the order, starting with behavioral theory (D&K) and then proposing an experimental test for empirical relevance. Moreover, some previous works (e.g., Fehr, Kremhelmer, and Schmidt 2008, p. 1277) suggest that formal models of intentions-based reciprocity make sense, *but* are difficult to apply. We agree as regards the make-sense part, but somewhat disagree on the difficult-to-apply part and hope this paper will serve as proof-of-concept.

Section II develops a simple hold-up framework which embeds game forms corresponding to Examples 1 and 2, and derives predictions for selfish players. Section III elucidates the general economic relevance of vengeance, reviewing contributions which are not explicitly focused on hold-up, but which indicate that vengeance is a powerful human motive (including experiments documenting how the presence of punishment options may have a dramatic impact on strategic interaction). Section IV develops theory, Section V reports experimental results, Section VI concludes.

II. A SIMPLE HOLD-UP MODEL

The two examples in Section I can be embedded in a simple hold-up framework: Two players, 1 and 2, have the opportunity to trade a unit of some investment good for which they have values v_1 and v_2, respectively. Assume that, prior to any moves, an implicit contract is formed under which the price of the investment good is p. Assume that investment is observable to both parties, but not verifiable; each party can observe the investment decision, but it cannot be explicitly contracted upon ex ante. Furthermore, assume that $v_1 = 0$, so the investment good has no residual value to 1.

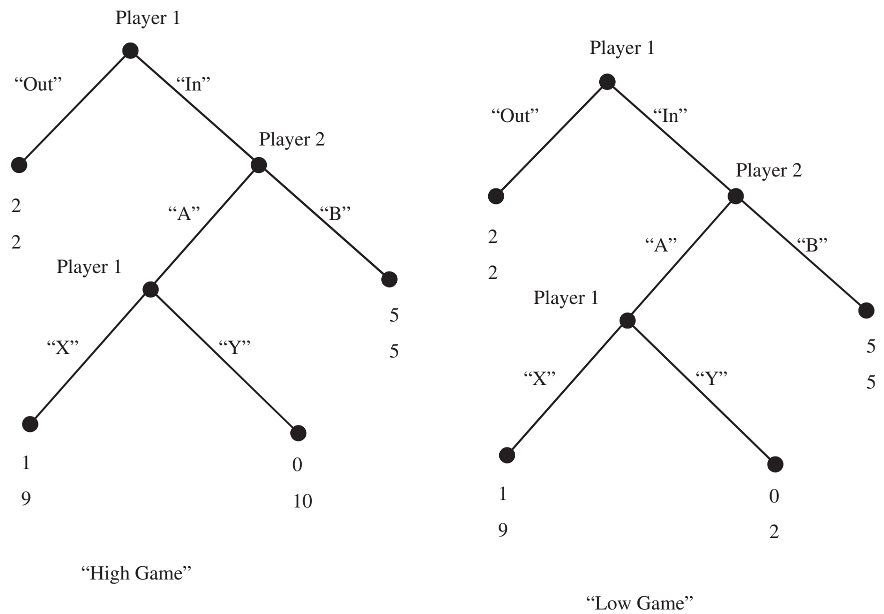
First, 1 chooses whether to invest. If not, then the players receive their reservation values of zero. Investment costs c. Upon observing 1's investment, 2 must decide whether to honor the implicit contract. If 2 does so, 1 receives a payoff of $p - c$ and 2 receives $v_2 - p$. If 2 chooses not to honor the implicit contract, he makes a take-it-or-leave-it offer t to 1. If 1 accepts this offer he receives a payoff of $t - c$ and 2 receives $v_2 - t$. If 1 does not accept then he must cover the investment cost so his payoff is $-c$ and 2 receives $(1 - \alpha)v_2$, where $\alpha \in [0, 1]$ is a measure of the residual control right to the investment held by 1. Full control over the investment proceeds by 1 is captured by $\alpha = 1$, while $\alpha = 0$ captures full control by 2. We assume that $v_2 > p > c > t > 0$ so investment is potentially beneficial to both parties. The two examples can now be accommodated by setting $(v_2, p, c, t) = (8, 5, 2, 1)$, where the numbers are measured in thousands of dollars. In Example 1, $\alpha = 1$ because the artist retains control over the painting. In Example 2, $\alpha = 0$ because the student retains control over his human capital.

Figure 1 recasts Examples 1 and 2 as extensive game forms with monetary payoffs. Relative to the examples, we have normalized the payoffs (dividing by 1000, then adding constants of 2 to each payoff entry) and simplified choice labels. The games represent stylized hold-up problems, with different underlying assumptions regarding residual rights of control. The Low-game corresponds to the case where Player 1 retains control after investment (as in Example 1), while in the High-game Player 2 retains these rights (as in Example 2).

The backward induction solution for selfish players is $((Out, X), A)$ in each game, with resulting inefficient payoffs (2, 2). If the players are selfish, the hold-up problem is thus equally severe in each case. In Section IV, we will see that incorporating negative reciprocity will change this conclusion. However, before we proceed in behavioral directions we wish to point out an important issue which clarifies the connection between our exercise and the traditional hold-up literature.

Examples 1 and 2, and so the Low-game and the High-game, impose a special structure on renegotiation: player 2 makes a take-it-or-leave-it-offer to player 1 (20% of the originally agreed upon price). From the viewpoint of traditional hold-up theory this is a rather special case, and as noted by Che and Sákovics (2008, Remark 1) in their recent survey of the hold-up literature the "effects of alternative ownership structures may depend on the bargaining solution assumed." If, for example, players engaged in Nash bargaining, then the nature of a renegotiated deal would depend on the players' payoffs should they not strike a deal (the "threat points"), which in turn would depend on residual rights of control. As players anticipate bargaining outcomes, this could affect the likelihood

FIGURE 1
Two Hold-Up Games



of hold-up for reasons other than negative reciprocity. The purpose of our paper is not to comment on how to best model renegotiation, however. Nor do we want any bargaining confound as we attempt to isolate the effects of negative reciprocity on hold-up as residual rights of control shift. We therefore, as seen in the Low-game and the High-game, assume a bargaining institution the outcome of which under selfish preferences is invariant to the assignment of residual rights of control.

III. THE CASE FOR NEGATIVE RECIPROCITY

That vengefulness can be a powerful motive is reflected in popular culture which is filled with stories of good (and sometimes bad) guys who get even. Our title makes reference to the 1995 movie *Die Hard: With a Vengeance*, starring Bruce Willis. Another example (disturbing and deeply touching) can be found in Cormac McCarthy's 2006 novel *The Road*, when the protagonist father decides to strip a thief of all he has (pp. 215–217). "I'm going to leave you the way you left us."

Real-world evidence is abundant too. Interview studies with business leaders, including that of Bewley (1999), conclude that an important reason for downward wage rigidity in times of recession is the perception that wage cuts would be perceived by employees as unkind measures to which the reaction would be less conscientious on-the-job effort.[4] The economic consequence may be nationwide involuntary unemployment, as firms prefer layoffs to wage cuts if the fired personnel cannot strike back. Firing personnel can have negative-reciprocity repercussions too though, as fired personnel have been known to engage in sabotage such as messing up computer systems before they leave their office.[5]

Businessmen can be vengeful too. Donald Trump, in his recent book *Think Big & Kick Ass* (Trump and Zanker 2007) which teaches how to be successful in business and life beyond, devotes an entire chapter to the importance of revenge. Part of the message reflects repeated-game or reputation concerns, but part is clearly reflecting an innate joy of getting even. The following passage illustrates (p. 198):

> Most business writers won't be so blunt and honest with you about getting even. They know it's the truth,

4. See Dufwenberg and Kirchsteiger (2000) for more references and a theoretical account.

5. This last example is taken from Sobel (2005, p. 393).

> but won't tell you because they want people to think of them as a "nice person." I don't like to mince words. When you are wronged and do nothing about it, you aren't "nice" you're a schmuck. That is why I say when you are wronged, go after the those people, because it's a good feeling . . . I love it.

Of course, the hold-up literature is concerned with interaction between businessmen (say CEOs or self-employed independent contractors); Trump's advice underscores the relevance of negative reciprocity in this connection.

Further documentation of the relevance of vengeful motivation to human decision making comes from studies of experimental games. Lopsided offers get rejected in ultimatum games, costly punishment options are exercised in public goods contexts, and low wage offers are met with low effort in wage-effort gift-exchange games, for example. We refer to Fehr and Gächter (2000a) for a discussion of much of the evidence and its interpretation with respect to reciprocity. Particularly telling evidence includes Bolton and Zwick's (1995) comparison of ultimatum games and impunity games (which look like ultimatum games except the responder can only reject the part of a proposal meant for him, so there cannot be punishment), and several studies that compare various games with and without added punishment options (e.g., Fehr and Gächter 2000b, 2002; Falk, Fehr, and Fischbacher 2005).

Theorists have responded to the evidence by developing formal models of reciprocity. Rabin (1993) is a pioneer; other papers include D&K, Falk and Fischbacher (2006), Segal and Sobel (2007), and Cox, Friedman and Sadiraj (2008).[6] There is also a literature which establishes that negative reciprocity preferences can be given evolutionary foundations. The key idea is that genes which instill an inclination for costly revenge may make others back off, which may increase one's rewards and so one's fitness. See Guth and Yaari (1992), Huck and Oechssler (1995), and Friedman and Singh (2009) for different approaches.

Reciprocity has two faces; negative reciprocity involves vengeance, while positive reciprocity involves rewarding kindness. Most previous discussions of reciprocity treat both forms alongside, but we focus on negative reciprocity. Our intuitions concerning Examples 1 and 2 involved retribution only, and our goal is to see how incorporating a taste for vengeance changes conclusions relative to a traditional hold-up analysis. Positive reciprocity may be an important human motivation,[7] but so may a host of other concerns like anxiety, guilt, disappointment, regret, fear, hope, altruism, and inequity aversion. In order to highlight the effect of negative reciprocity, we abstract away from all these other sentiments, including positive reciprocity.

IV. THEORY

Rabin (1993) highlights key qualitative aspects of reciprocity in games. However, as he notes himself (p. 1296), his normal form construction is inadequate for applications to games with a nontrivial dynamic structure (like our Low- and High-games) which require a description of how kindness conceptions get updated through a game tree. D&K therefore develop a model for extensive form games, which considers both positive and negative reciprocity. As explained in Section III, we wish to focus solely on negative reciprocity and propose a modification of D&K's theory which achieves this. We first recall D&K's theory (sketchily; for the unabridged version with its formal details we refer to D&K), then introduce the negative-reciprocity-only modification, and finally apply this theory to our games.

A. *Dufwenberg and Kirchsteiger (2004)*

D&K consider finite multistage games with observed actions and without nature (this includes our Low- and High-games). *Ecco* player i's utility:

$$(1)\quad u_i = \underbrace{\pi_i}_{\text{material payoff}} + \underbrace{\Sigma_j(\theta_{ij} * \kappa_{ij} * \lambda_{iji})}_{\text{reciprocity payoff}}$$

Player i's material payoff π_i simply reflects his dollar earnings. $\theta_{ij} * \kappa_{ij} * \lambda_{iji}$ is i's reciprocity payoff with respect to player j: $\theta_{ij} \geq 0$ is a parameter reflecting i's sensitivity to reciprocity (D&K use Y_{ij} instead of θ_{ij}). κ_{ij} is i's kindness to j; this term is negative (positive) if i is unkind (kind) to j. λ_{iji} describes how kind

6. We focus on approaches that link reciprocity to intentions; we do not emphasize the complementary approach where players care about forms of inequity of distributions. Fehr and Schmidt (2003) and Sobel (2005) also discuss such models.

7. Some experiments suggest that negative reciprocity is more important than positive reciprocity though. See Charness (2004), Charness and Rabin (2002), Offerman (2002).

i perceives j to be; in analogy with kindness, this term is negative or positive. Reciprocity is captured as, *ceteris paribus*, i wants to match the sign of κ_{ij} and λ_{iji}, to make the product $\kappa_{ij} * \lambda_{iji}$ positive.

κ_{ij} is actually a real-valued function of i's strategy and "first-order beliefs" about other players' strategies; κ_{ij} is the difference between the material payoff i believes j gets and the average of the maximum and minimum material payoffs that i believes j could have gotten had i chosen differently.[8] Perceived kindness λ_{iji} is defined analogously, except that its arguments—i's first-order belief about j's strategy and i's "second-order beliefs" about j's first-order beliefs—appear one level higher in i's belief hierarchy.

This *Reader's Digest* presentation of D&K's theory warrants a few more comments: D&K define a solution concept called *sequential reciprocity equilibrium* (*SRE*). The SRE concept imposes that players optimize at all histories, for beliefs which are "correct." This involves that strategies and beliefs at any history reflect probability 1 choices along the path that lead to that history; this is how D&K model intentionality and updating of kindness as play proceeds. As i's utility u_i includes beliefs as arguments, D&K's model fits the framework of psychological games (Geanakoplos, Pearce, and Stacchetti 1989; Battigalli and Dufwenberg 2009).

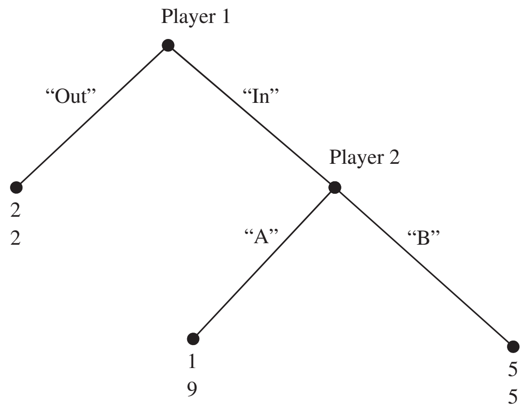
B. Vengeance Equilibrium

We now modify D&K's theory to reflect negative reciprocity only. Rather than operating with perceived kindness function λ_{iji}, we simply define $\lambda^{-}_{iji} := \min\{\lambda_{iji}, 0\}$. Rather than maximizing u_i, player i maximizes his utility v_i (v for "vengeance") defined by:

$$(2) \quad v_i = \underbrace{\pi_i}_{\text{material payoff}} + \underbrace{\Sigma_j(\theta_{ij} * \kappa_{ij} * \lambda^{-}_{iji})}_{\text{vengeance payoff}}$$

Define a *Vengeance Equilibrium* (*VE*) just like a SRE, except that i's utility is v_i rather than u_i. Note two properties of the VE: First, a VE exists for any game within the considered class; D&K's proof applies directly, *mutatis mutandis*. Second, neither SRE nor VE is a refinement of

FIGURE 2
VE versus SRE



the other concept. To see this, consider the game in Figure 2.

It can be verified that if θ_{12} and θ_{21} are high enough then the unique SRE is (In, B), while the unique VE is (Out, A).

C. Hold-up with Vengeance

We are now ready to solve our two hold-up games for their VEs, paying special attention to the cases where θ_{12} and θ_{21} are large numbers so that concern for getting even is important. We now start with the High-game where the prediction is unambiguous:

OBSERVATION 1. *Regardless of* θ_{12} *and* θ_{21}, *the unique* VE *of the High-game is* $((Out, X), A)$.

To verify Observation 1, check incentives through the game: *At the root*, given equilibrium beliefs, player 1 believes player 2's kindness is zero ($\lambda_{121} = 0$), as 2 does not affect material payoff if 1 chooses *Out*. Therefore, at the root, 1 will maximize his material payoff. He would obtain a payoff of 1 choosing *In* (because play would proceed A then X), while Out gives him $2 > 1$.... *Next, look at 2's node*: Given (updated) equilibrium beliefs, $\lambda_{212} > 0$ so $\lambda^{-}_{212} = 0$, so 2 will maximize his/her material payoff. As player 1 chooses X if given the chance, 2's best response is A.... *Finally, after history* (In, A): 1's reciprocity and material payoffs are aligned; 1's material reward is higher if he/she chooses X rather than Y, and because 2's choice A is unkind 1's reciprocity

8. More precisely, the calculation of the minimum is defined with respect to what D&K call i's "efficient" strategies. See D&K for details. The qualification is irrelevant for analyzing our High- and Low-games, so we neglect this somewhat complicated issue here.

payoff is also maximized by choice X which minimizes 2's material payoff.

Observation 1 shows that the High-game embodies inefficiency because of hold-up whether or not players are vengeful. Observation 2, by contrast, shows that vengefulness may mitigate hold-up in the Low-game:

OBSERVATION 2. *If θ_{12} and θ_{21} are large enough, there are two VE's of the Low-game:*

(i) $((In, Y), B)$ *and*
(ii) $((Out, Y), A)$.

Case (i) captures our intuitions regarding Example 1 from Section I: *At the root*, $\lambda_{121} > 0$, that is, $\lambda^{-}_{121} = 0$, so 1 will maximize his material payoff. He is happy with his *In* choice, because $5 > 2$.... *At 2's node*, he/she chooses B. To see why, note that given equilibrium beliefs, $\lambda_{212} > 0$ so $\lambda^{-}_{212} = 0$. Thus 2 will maximize her material payoffs, and by choosing B he/she gets 5 rather than 2.... *Finally, after history* (In, A), player 1 is motivated by negative reciprocity to choose Y; at that history $\lambda^{-}_{121} = \lambda_{121} < 0$, so with θ_{12} high enough A will minimize B's material payoff.

Take a moment to reflect on the following "irony" concerning the VE of Case (i) as compared to the unique VE of the High-game: The Low-game differs from the High-game in that a player's payoffs goes *down* (a 10 becomes a 2), but in VE both players' material payoffs go up.

However, the strategy profile $((In, Y), B)$ is not the only VE-possibility in the Low-game. Case (ii) adds an intriguing, gloomy possibility. The strategy profile $((Out, Y), A)$ is also a VE. We like to think of this strategy profile as a "miserable" *VE*. This time, we discuss what is going on starting below the root.

In the subgame starting at 2's node (off the equilibrium path) the two players indulge in being unkind to one another. Given (updated) equilibrium beliefs, each player perceives the other as unkind, and strikes back to minimize the coplayer's payoff. For player 1, the argument is just like in Case (i). For player 2, the behavior could be thought of as "reciprocation-in-anticipation," as 1 makes the choice that hurts 2 *after* 2 makes the choice that hurts 1. This is allowed by the theory; given (updated) equilibrium beliefs, at 2's node, he/she believes that 1 believes he will willingly minimize her material payoffs; $\lambda^{-}_{212} = \lambda_{212} < 0$, so 2 is happy to minimize 1's material payoff by choosing A.... *Finally, going back to the root*, given equilibrium beliefs, 1 realizes that strategies (In, Y) and (Out, Y) both give player 2 a material payoff of 2, so κ_{12} is the same for both strategies. As the material payoff from choosing *Out* is higher, this breaks what in terms of reciprocity payoff is a tie in favor of (*Out*, Y).

It is natural to wonder whether the conclusions we have obtained here for our High- and Low-games are robust with respect to changes in the underlying parameters. The answer is *yes*: in Appendix B, we return to the broader hold-up framework presented at the start of Section II and show that Observations 1 and 2(i) have counterparts in this more general setting.

V. EXPERIMENT

Our work falls in the category of "applied theory." It should primarily be evaluated according to how intriguing and novel is our examination of connections between vengefulness and residual rights of control for hold-up. Nevertheless, theoretical stories, however intriguing, gain some value if they also possess empirical relevance. Against this background, we ran an experiment to test the predictions of section IV. This section explains what we found.

A. Testable Predictions

As seen in Section IV.C, incorporating negative reciprocity does not change conclusions regarding the severity of hold-up in the High-game, as $((Out, X), A)$ is its unique VE regardless of θ_{12} and θ_{21} (Observation 1). However, in the Low-game, if θ_{12} and θ_{21} are high enough, *every choice may flip*, in the sense that $((In, Y), B)$ is a VE (Observation 2, Case (i)). It is thus natural to test whether, on balance, the choices *In*, Y, and B are more common than *Out*, X, and A in the Low-game than in the High-game.

One potential counter-force to this prediction is the additional VE pointed out in Case (ii) of Observation 2: $((Out, Y), A)$. The existence of this VE increases the possibility that we may not be able to reject the hypotheses that the behavior in the two games, at each decision node, is the same. In other words, if our subjects are attracted to the miserable equilibrium behavior across treatments will look similar even if subjects are vengeful. On the other hand, if we do get rejection it will support the idea that negative reciprocity can mitigate hold-up mainly in cases where the investing party maintains residual rights of control, as well as the empirical relevance of the $((In, Y), B)$ VE in the Low-game.

B. Procedures

All lab sessions were conducted in the Economic Science Laboratory at the University of Arizona. Subjects were undergraduate students recruited via E-mail from our online subject database. Twelve subjects participated in each session, six subjects were randomly assigned to the Low-treatment and the rest were assigned to the High-treatment. In the Low-treatment, subjects played the Low-game. In the High-treatment, subjects played the High-game. Five sessions were conducted.

Upon arriving at the lab, subjects were checked-in and randomly assigned to individual computer carrels (the computer software was programmed using ECONPORT). Half of the subjects found high treatment instructions when they arrived at their carrel and the other half found low treatment instructions. The instructions are available in Appendix A.

After reading the instructions on their own, the experimenter answered any questions subjects had privately. The same experimenter was present at each of the sessions. After all the questions had been answered the experiment began. Subjects were randomly assigned by the computer to be either "Player 1" or "Player 2" for the duration of the experiment.

Our games are complicated enough that we thought it wise to let each subject play several times, to allow for some learning and possible gravitation toward equilibrium. At the same time, to maintain the one-shot nature of the interaction presumed by the theory and to avoid creating a repeated game, we employed a random-matching format. In each treatment, subjects played the game for five rounds. In each round, each subject was randomly and anonymously matched with one of the other players in their group. In each round, choices were recorded using the ECONPORT software, while play proceeded sequentially through the game. All of this information was given to subjects in the instructions.

In each game, subjects earned dollars corresponding to the payoff numbers described in our two games. Subjects recorded their per-round payoff on a summary sheet and were privately paid their cumulative profit at the end of the session. No exchange rate was used. Sessions typically lasted about 30 min. No subject participated in more than one session.

C. Results

Our model of vengeance equilibrium predicts behavior conditional upon each history and on each player's (updated) beliefs, given that players attach a sufficiently large utility weight to vengeance. The theoretical analysis suggests that subjects' choices would gravitate toward strategy profile $((Out, X), A)$ in the High-game, and toward strategy profile $((In, Y), B)$ in the Low-game to the extent that the Case (i) VE is relevant. Comparing these two strategy profiles, we get the across-treatment prediction that *every choice should flip*.

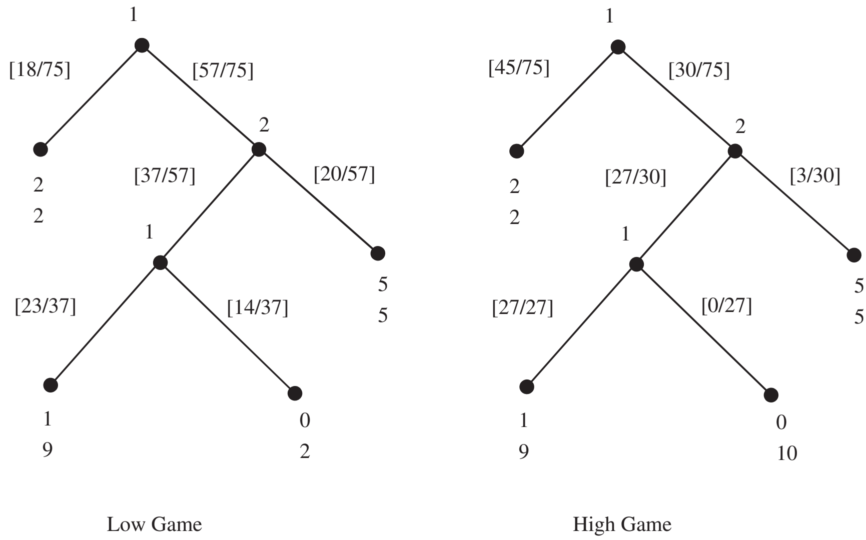
Figure 3 presents experimental results which aggregate behavior at each history, across subjects and rounds, in order to compare the predictions of our equilibrium analysis with observed behavior. At each history the difference in observed choice frequencies is consistent with our hypotheses: a one-sided Fisher exact test for difference in observed choice frequencies shows significant differences at the 0.001 or greater level for each of the three histories.

Our second analysis uses the average number of *In*, *B*, and *Y* observations for each group at every session. This gives us a total of five independent observations at each of the three information sets for each treatment. As the number of independent observations is small with this approach, we employ nonparametric statistical tests.

At each of the three information sets, we use the Mann-Whitney test (see Siegel and Castellan 1988). We conduct three separate tests to determine whether the two treatments have the same mean percentage of *In*, *B*, and *Y* moves. For each test, the alternative hypothesis is that the mean percentage is higher for the Low-treatment. This gives us three directional tests for our research hypotheses.

We present the results from the different stages of the game in "backward" order, starting with the stage after Player 1 has chosen *In* and Player 2 has chosen *A*. Player 1 now has the choice of *X* or *Y*. Recall that the key difference between the two treatments as regards material payoff occurs at this stage: By choosing *Y* rather than *X* player 1 gives up a material payoff of 1 in both treatments, while player 2's material payoff is thereby reduced (from 9 to 2) in the Low-treatment and increased (from 9 to 10)

FIGURE 3
Summary of Experimental Results



in the High-treatment. Our research hypothesis at this stage is that *the mean percentage of Y choices is higher in the Low-game than in the High-game.* Table 1 records mean percentage data for the five independent sessions.

A casual look at the data confirms the willingness of subjects to engage in costly punishing once play had reached the third stage. Under the null, the probability of observing a sample as extreme as this one is 0.0027. We therefore clearly reject the associated null. This willingness to punish, even to the detriment of one's own payoff, after player 2 chose action *A* supports the idea of a negative reciprocity motivation.

The second stage of the game is when player 2 chooses *B* or *A*, following player 1's choice of *In*. Our research hypothesis at this stage, following the discussion in Sections IV.C and V.A, is that *the mean percentage of B choices is higher in the Low-game than in the High-game.* Table 2 records mean percentage data for the five independent sessions.

Under the null, the probability of observing this sample, or one as extreme, as this one is 0.0362. We thus reject the null. In other words, we find support for the idea that conditional on Player 1 playing *In*, the efficient equal-split is more likely in the Low-game than the High-game.

At the first stage player 1 chooses whether to trust player 2; *In* or *Out*. In the Low-game, player 1 knows he has a punishment mechanism if player 2 chooses *A*. Our research hypothesis, following the discussion in Sections IV.C and V.A, is that *the mean percentage of In choices is higher in the Low-game than in the High-game.* Table 3 records mean percentage data for the five independent sessions.

Under the null hypothesis that the two samples come from the same distribution, the probability of observing this outcome, or one that is more extreme, is 0.0102. We therefore reject

TABLE 1
Final Stage Choices (Fraction *Y*)

Treatment	Session 1	Session 2	Session 3	Session 4	Session 5
Low	0.1250	0.4444	0.2500	0.5714	0.6000
High	0.000	0.0000	0.0000	0.0000	0.0000

TABLE 2
Second Stage Choices (Fraction *B*)

Treatment	Session 1	Session 2	Session 3	Session 4	Session 5
Low	0.0000	0.2500	0.4286	0.2222	0.6000
High	0.0000	0.1111	0.1429	0.2000	0.0000

TABLE 3
Root Choices (Fraction *In*)

Treatment	Session 1	Session 2	Session 3	Session 4	Session 5
Low	0.5333	0.8000	0.9333	0.6000	0.9333
High	0.2667	0.6000	0.4667	0.3333	0.3333

the null. It is plain that in all five sessions the mean percentage of *In* choices was higher in the Low-treatment than in the High-treatment.

We noted in Section III that to the extent that the miserable VE described there would have been relevant to the Low-game, negative reciprocity could have been an important motivational force even if there would not have been much of a difference in the nature of play between the High-game and the Low-game. In light of the data, this point now seems moot. All in all, we take the support for our research hypotheses as reinforcing the idea that negative reciprocity can mitigate hold-up mainly in cases where the investing party maintains the residual rights of control.

VI. CONCLUDING REMARKS

The back cover of the *JPE* once recalled a hold-up story about a rich woman in Savannah where, between the lines, we see negative reciprocity at work[9]:

> Some years ago she ordered a pair of iron gates for her house. They were designed and built especially for her. But when they were delivered she pitched a fit, said they were horrible, said they were filth. “Take them away,” she said, “I never want to see them again!” Then she tore up the bill, which was for $1,400—a fair amount of money in those days.
>
> The foundry took the gates back, but didn’t know what to do with them . . . there wasn’t much demand for a pair of ornamental gates exactly that size. The only thing they could do was sell the iron for its scrap value. So they cut the price from $1,400 to $190. Naturally, the following day the woman sent a man over to the foundry with $190, and today those gates are hanging on her gateposts where they were originally designed to go.

The story may seem puzzling. Why would the woman send a man to the foundry rather than just make a take-it-or-leave it offer herself? Part of the answer may be that she feared a counter-offer, but another part is that she might otherwise irritate the foundry’s owner who may retaliate by refusing to sell her the gate. On this interpretation, we thus have a situation where a proper understanding of an economic outcome involves reference to negative reciprocity. And if we modify the situation to make the foundry less naive, that is, so that they could see through the woman’s ploy, the situation would structurally resemble our Example 1, or our Low-game.

Classical hold-up theory typically assumes that the involved parties selfishly maximize own income. We have argued that this perspective may be too limited; negative reciprocity may plausibly play a role too. Injured parties may have an inclination to strike back if they are treated badly (even if this is costly), and if this is anticipated the problems because of hold-up are mitigated. We have shown, however, that it would be premature to draw the blanket conclusion that hold-up is not a serious concern. Rather, this depends in predictable ways on details of the situation. Namely, hold-up is a less serious concern if the investing party retains residual rights of control than if the other party does. This conclusion is supported by a D&K-based theory of negative reciprocity which we apply to two hold-up games (which vary the residual right of control), and through a related experimental test.

9. See *Journal of Political Economy* 107(1), February 1999. The excerpt is from John Berendt’s 1994 novel *Midnight in the Garden of Good and Evil*. It was suggested to the *JPE* by Oliver Hart, and to us by Tore Ellingsen.

Under certain conditions, our conclusions accord well with those that come out of traditional hold-up analysis. If one assumes (as we did not) that renegotiations take the form of Nash bargaining, then if the noninvesting party holds the residual rights of control he has a favorable threat point which however serves him badly because the other party may shun a deal for fear of hold-up. Assuming Nash bargaining and incorporating negative reciprocity may thus have similar consequences, a fascinating insight because Nash's bargaining theory is not motivated with reference to vengeance. To properly gauge the seriousness of hold-up one may independently assess how details of an economic situation affect bargaining power as well as the incentives for decision makers inclined to get even.

The insights of this paper may matter not only to theorists but also to practitioners who conduct industry analysis with the aim of assessing optimal responses to market conditions. Evaluating hold-up problems (upstream and downstream) may be important to schemata such as "Porter's Five Forces." See Porter (1980) or, for more recent textbook guidance that devotes considerable attention to hold-up, Besanko et al. (2010; pp. 140–146).[10] We have not seen any reference in strategy textbooks to negative reciprocity, and its link to residual rights of control.

APPENDIX A

Instructions (High Game)

Instructions. This is an experiment in strategic decision making. If you read these instructions carefully and pay attention to the Experimenter, you have the potential to earn a considerable amount of money. The decisions that are made will affect the payments of everyone involved in the game.

You will be playing 5 rounds of a simple game that will be described below. At the beginning of the experiment you will be randomly selected by the computer to play the role Player 1 or Player 2. At the beginning of each round you shall be randomly matched with another person in the room to play the game, but you will maintain the same player role in each round. At the end of the 5 rounds, you will be privately paid your total earnings.

The Game:

- Player 1 goes first and must decide whether to play "In" or "Out."

 – If "Out" is selected, the game will be over with both players receiving a payment of 2 each. If "In" is selected, it will be Player 2's turn to make a decision.

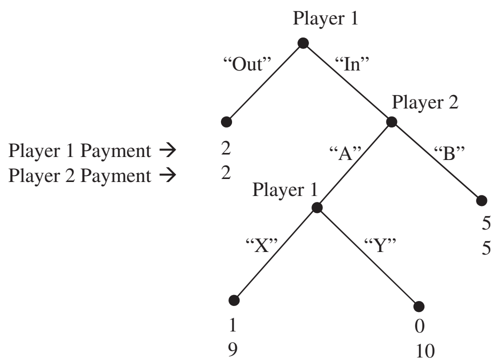
- If Player 1 chooses to play "In," Player 2 must decide whether to play "A" or "B."

 – If "B" is selected, the game will be over with both players receiving a payment of 5 each. If "A" is selected, it will again be Player 1's turn to make a decision.

- If Player 2 plays "A," Player 1 must decide to play "X" or "Y."

 – If "X" is selected, Player 1 receives a payment of 1 and Player 2 receives a payment of 9. If "Y" is selected, Player 1 receives a payment of 0 and Player 2 receives a payment of 10.

The following picture may help. Player 1's payments are listed on the top.



This completes the description of the game. Do you have any questions?

Instructions (Low Game)

Instructions. This is an experiment in strategic decision making. If you read these instructions carefully and pay attention to the Experimenter, you have the potential to earn a considerable amount of money. The decisions that are made will affect the payments of everyone involved in the game.

You will be playing 5 rounds of a simple game that will be described below. At the beginning of the experiment you will be randomly selected by the computer to play the role Player 1 or Player 2. At the beginning of each round you shall be randomly matched with another person in the room to play the game, but you will maintain the same player role in each round. At the end of the 5 rounds, you will be privately paid your total earnings.

The Game:

- Player 1 goes first and must decide whether to play "In" or "Out."

 – If "Out" is selected, the game will be over with both players receiving a payment of 2 each. If "In" is selected, it will be Player 2's turn to make a decision.

- If Player 1 chooses to play "In," Player 2 must decide whether to play "A" or "B."

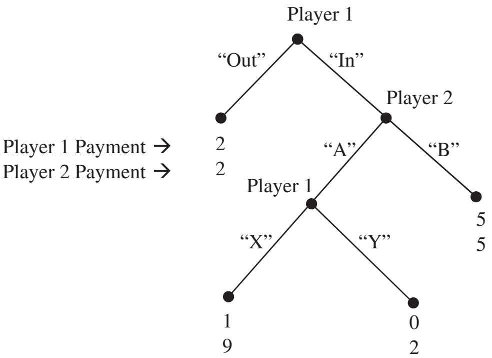
 – If "B" is selected, the game will be over with both players receiving a payment of 5 each. If "A" is selected, it will again be Player 1's turn to make a decision.

10. See also Besanko et al's scorecard items that concern "relation-specific investments" in the appendix to chapter 12, and question 10 in their chapter 5, which inspired our Examples 1 and 2 although we added the focus on residual rights of control.

• If Player 2 plays "A," Player 1 must decide to play "X" or "Y."

– If "X" is selected, Player 1 receives a payment of 1 and Player 2 receives a payment of 9. If "Y" is selected, Player 1 receives a payment of 0 and Player 2 receives a payment of 2.

The following picture may help. Player 1's payments are listed on the top.



This completes the description of the game. Do you have any questions?

APPENDIX B

In this section, drawing on D&K and the VE concept defined in Section IV.B, we extend Observations 1 and 2(i) to the generalized hold-up model of section 2 characterized by $v_2 > p > c > t > v_1 = 0$ and $\alpha \in [0, 1]$.

PROPOSITION 1. *Fix* $\theta_{12} \geq 0$ *and* $\theta_{21} \geq 0$. *If* $\theta_{12} = 0$, *then* $[(Out,\ X),\ A]$ *is the unique VE of the problem for any* $\alpha \in [0, 1]$. *If* $\theta_{12} > 0$, *then* $[(Out,\ X),\ A]$ *is the unique VE of the generalized hold-up problem if* $\alpha < \frac{t}{v_2} + \frac{2t}{\theta_{12} v_2 (p-t)}$.

Proof. After the history (In, A), 1 always perceives a choice of A by 2 as unkind (because $\lambda^{-}_{121} < 0$ for all second order beliefs). It follows that 1 chooses X over Y if $t > \theta_{12}(\kappa_{12}(Y, \cdot) - \kappa_{12}(X, \cdot))\lambda^{-}_{121}$, which is satisfied for any $\alpha \in [0, 1]$ if $\theta_{12} = 0$ or for $\alpha < \frac{t}{v_2} + \frac{2t}{\theta_{12} v_2 (p-t)}$ if $\theta_{12} > 0$. If 2 believes 1 will choose X, then 2 will choose A. This is because 2 believes 1's strategy (In, X) is kind (i.e., $\lambda_{212} > 0$), so $\lambda^{-}_{212} = 0$. Therefore only material payoff matters to 2. As $v_2 - c > v_2 - p$, both 2's utility and material payoffs are higher from A given that 1 chooses X. A can best respond to this at the root with *Out*; this maximizes 1's material payoff while the reciprocity payoff is zero given equilibrium beliefs (because $\lambda_{121} = 0$ given that 1 chooses *Out*). To finally infer uniqueness, note that given 1's belief that 2 will choose A we have $\lambda^{-}_{121} < 0$ once 2 moves. As *Out* minimizes 1's kindness to 2, *Out* in fact maximizes both 1's material and vengeance payoffs. ■

PROPOSITION 2. *For every* $\theta_{21} \geq 0$ *and* $\alpha \in (\frac{p}{v}, 1]$, *there exists a* $\overline{\theta}_{12} > 0$ *such that for* $\theta_{12} \geq \overline{\theta}_{12}$ *we have* $((In, Y), B)$ *is a VE of the generalized hold-up game.*

Proof. Impose equilibrium beliefs. Then after the history (In, A), 1 always perceives a choice of A by 2 as unkind (because $\lambda^{-}_{121} = -\frac{p}{2} < 0$. It follows that 1 chooses Y over X only if $t \leq \theta_{12}(\kappa_{12}(Y, \cdot) - \kappa_{12}(X, \cdot))\lambda^{-}_{121}$, which is satisfied if $\theta_{12} \geq \frac{2t}{p(\alpha v_2 - t)} \equiv \overline{\theta}_{12}$. Equilibrium beliefs require that 2 believes 1 will choose Y and that 2 believes 1 believes 2 will choose B, thus 2 believes (In, Y) is kind (i.e., $\lambda_{212} = 0$), so $\lambda^{-}_{212} = 0$. Therefore only material payoff matters. As $v_2 - p > (1 - \alpha)v_2$, both 2's utility and material payoffs are higher from B given that 1 chooses X. Finally, if 1 believes 2 will choose B, then $\lambda^{-}_{121} = 0$, and 1 will choose *In* to maximize his material payoff. ■

REFERENCES

Battigalli, P., and M. Dufwenberg. "Dynamic Psychological Games." *Journal of Economic Theory*, 144, 2009, 1–35.

Besanko, D., D. Dranove, M. Shanley, and S. Schaefer. *Economics of Strategy*. 5th ed. Hoboken, NJ: John Wiley & Sons, 2010.

Bewley, T. *Why Wages Don't Fall During a Recession*. Cambridge, MA: Harvard University Press, 1999.

Bolton, G.E., R. Zwick "Anonymity versus Punishment in Ultimatum Bargaining." *Games & Economic Behavior*, 10, 1995, 95–121.

Charness, G. "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics*, 22, 2004, 665–88.

Charness, G., and M. Rabin. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117, 2002, 817–69.

Che, Y., and J. Sákovics. "Hold-Up Problem," in *The New Palgrave Dictionary of Economics*, edited by S. N. Durlauf and L. E. Blume. 2nd ed. New York: Palgrave Macmillan, 2008.

Cox, J., D. Friedman, and V. Sadiraj. "Revealed Altruism." *Econometrica*, 76, 2008, 31–69.

Dufwenberg, M., and G. Kirchsteiger. "Reciprocity & Wage Undercutting." *European Economic Review*, 44, 2000, 1069–78.

———. "A Theory of Sequential Reciprocity." *Games & Economic Behavior*, 47, 2004, 268–98.

Ellingsen, T., and M. Johannesson. "Promises, Threats, and Fairness." *The Economic Journal*, 114, 2004a, 397–420.

———. "Is There a Hold-up Problem?" *Scandinavian Journal of Economics*, 106, 2004b, 475–94.

Falk, A., and U. Fischbacher. "A Theory of Reciprocity." *Games & Economic Behavior*, 54, 2006, 293–315.

Falk, A., E. Fehr, and U. Fischbacher. "Driving Forces behind Informal Sanctions." *Econometrica*, 73, 2005, 2017–30.

Fehr, E., and S. Gächter. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14, 2000a, 159–81.

———. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90, 2000b, 980–94.

———. "Altruistic Punishment in Humans." *Nature*, 415, 2002, 137–40.

Fehr, E., O. Hart, and C. Zehnder. "Contracts as Reference Points—Experimental Evidence." Mimeo, 2008.

Fehr, E., S. Kremhelmer, and K. Schmidt. "Fairness and the Optimal Allocation of Ownership Rights." *Economic Journal*, 118, 2008, 1262–84.

Fehr, E., and K. Schmidt. "Theories of Fairness & Reciprocity—Evidence & Economic Applications," in *Advances in Economics & Econometrics: 8th World*

Congress of the Econometric Society, edited by M. Dewatripont, L. Hansen, and S. Turnovsky. New York: Cambridge University Press, 2003.

Friedman, D., and N. Singh. "Vengefulness Evolves in Small Groups." *Games & Economic Behavior*, 66, 2009, 813–29.

Gantner, A., W. Guth, and M. Konigstein. "Equitable Choices in Bargaining Games with Joint Production." *Journal of Economic Behavior & Organization*, 46, 2001, 209–25.

Geanakopos, J., D. Pearce, and E. Staccetti. "Psychological Games and Sequential Rationality." *Games & Economic Behavior*, 1, 1989, 60–79.

Grossman, S., and G. Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94, 1986, 691–719.

Grout, P. "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach." *Econometrica*, 52, 1984, 449–60.

Güth, W., and M. Yaari. "An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game," in *Explaining Process and Change—Approaches to Evolutionary Economics*, edited by U. Witt. Ann Arbor: University of Michigan Press, 1992, 23–34.

Hackett, S. "Is Relational Exchange Possible in the Absence of Reputations and Repeated Contact?" *Journal of Law, Economics & organization*, 10, 1994, 360–89.

Hart, O. *Firm, Contracts, and Financial Structure*. Oxford: Clarendon Press, 1995.

———. "Hold-up, Asset Ownership, and Reference Points." *Quarterly Journal of Economics*, 123, 2008, 1–48.

Hart, O., and J. Moore. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98, 1990, 1119–58.

———. "Contracts as Reference Points." *Quarterly Journal of Economics*, 123, 2008, 1–48.

Huck, S., and J. Oechssler. "The Indirect Evolutionary Approach to Explaining Fair Allocations." *Games & Economic Behavior*, 28, 1995, 13–24.

Klein, B., R. Crawford, and A. Alchian. "Vertical Integration, Appropriable Rents, and the Comparative Contracting Process." *Journal of Law & Economics*, 21, 1978, 297–326.

Offerman, T. "Hurting Hurts More Than Helping Helps." *European Economic Review*, 46, 2002, 1423–37.

Porter, M. "Competitive Strategy—Techniques for Analyzing Industries and Competitors." First Free Press Edition. New York: Simon and Schuster, 1980.

Rabin, M. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83, 1993, 1281–302.

Segal, U., and J. Sobel. "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings." *Journal of Economic Theory*, 136, 2007, 197–216.

Siegel, C., and J. Castellan. *Nonparametric Statistics for the Behavioral Sciences*. 2nd ed. New York: McGraw-Hill, 1988.

von Siemens, F. "Bargaining under Incomplete Information, Fairness, and the Hold-Up Problem." *Journal of Economic Behavior & Organization*, 71, 2009, 486–94.

Sloof, R., J. Sonnemans, and H. Oosterbeek. "Does Making Specific Investments Unobservable Boost Investment Incentives?" *Journal of Economics & Management Strategy*, 16, 2007, 911–42.

Sobel, J. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature*, 43, 2005, 392–436.

Tirole, J. "Procurement and Renegotiation." *Journal of Political Economy*, 94, 1986, 235–59.

———. *Industrial Organization*. Cambridge, MA: MIT Press, 1988.

Trump, D., and B. Zanker. *Think Big & Kick Ass in Business and Life*. NYC: Harper Collins Publishers, 2007.

Williamson, O. *Markets and Hierarchies: Analysis and Antitrust Implications*. NYC: Free Press, 1975.

———. *The Economic Institutions of Capitalism*. NYC: Free Press, 1985.