Exp Econ (2011) 14: 84–109
DOI 10.1007/s10683-010-9258-0

Learning to respect property by refashioning theft into trade

Erik O. Kimbrough

Received: 6 October 2009 / Accepted: 23 September 2010 / Published online: 20 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Agent-based simulations and human-subject experiments explore the emergence of respect for property in a specialization and exchange economy with costless theft. Software agents, driven by reciprocity and hill-climbing heuristics and parameterized to replicate humans when property is exogenously protected, are employed to predict human behavior when property can be freely appropriated. Agents do not predict human behavior in a new set of experiments because subjects innovate, constructing a property convention of "mutual taking" in 5 out of the 6 experimental sessions that allows exchange to crowd out theft. When the same convention is made available to agents, they adopt it and again replicate human behavior. Property emerges as a social convention that exploits the capacity for reciprocity to sustain trade.

Keywords Property rights · Conventions · Experimental economics · Agent-based modeling

JEL Classification B41 · C63 · C92 · D51 · D83 · F10

Data, simulation code, and CSW experiment instructions available upon request. All simulations, data analysis, and figures were performed or created in R: A language and Environment for Statistical Computing (R Development Core Team 2010) with assistance from various contributed packages including: Bengtsson (2003), Berkelaar et al. (2008), Henningsen (2008), Neuwirth (2007), and Warnes et al. (2008).

Electronic supplementary material The online version of this article (doi:10.1007/s10683-010-9258-0) contains supplementary material, which is available to authorized users.

E.O. Kimbrough (✉)
Department of Economics (AE1), School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
e-mail: ekimbrough@gmail.com

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

1 Introduction

Property is the glue that holds together an economy based on exchange. For two agents to engage in mutually beneficial trade, it is assumed that each has a right to determine the fate of the goods and services being exchanged. Parties to an exchange rely on the presumption that their counterpart will neither renege on a partially completed transaction nor obviate the need for exchange by forcibly taking the desired goods. Such observations, while perhaps obvious, can be considered trivial only if one does not reflect on the fragility of property and the ease with which it can be violated by sufficiently motivated groups or individuals. The range of violations stretches from the historical predations of Viking marauders to lowly pickpockets in crowded subway stations, and property has always been precarious to a greater or lesser degree. How, then, does the respect for property emerge in order to facilitate specialization and trade? This paper seeks to answer that question with a combination of human-subject experiments and agent-based models. When subjects (or computerized agents) are placed in an environment with imperfect property protection, how do they come to respect one another's possessions in order to reap the gains from trade?

Kimbrough (forthcoming) develops an agent-based model of the two-good production and exchange economy of Crockett et al. (2009, hereafter CSW) to make concrete the dynamics and behavioral rules that guide subjects to discover (or fail to discover) specialization and trade. Hill-climbing and reciprocity heuristics are consistent with the varieties of human behavior in the original CSW environment and predict behavior under environmental variations with exogenous property enforcement. As Kimbrough argues, if agents' decision rules yield model output that is accurate in its depiction of human behavior in one environment, a *valid* model should predict human behavior in additional environments.[1] Hence this paper asks whether a calibrated agent-based model populated with heuristic-driven agents predicts human-subject behavior in a *new* set of experiments in which property is not exogenously protected.

This interplay between laboratory experiments and agent-based models offers an important method of testing behavioral explanations of economic outcomes because creating agents requires specifying detailed decision rules for which simulated interactions with an economic environment constitute predictions about human behavior. Many behavioral explanations from psychology, neuroscience and other disciplines can be compared by first formulating them as decision rules for computerized agents and then asking whether agents employing these rules in a given environment are able to predict human behavior in that same environment. Feedback from comparisons to additional human-subject data facilitates refinement of the decision model and may suggest additional experimental treatments necessary to settle disputes.[2]

[1] Arthur (1991) offers an early agent-based model that replicates human decisions in a simple, two-choice bandit experiment, but the degrees of freedom in programming a model make it relatively trivial (given enough time) to replicate human subject data. If a model is required also to make out-of-sample predictions under environmental variations, then the generality of the underlying decision rules can be better established. See Duffy (2001) and Arifovic and Ledyard (2004) for other implementations of this methodology.

[2] Recourse to additional data is essential to the validation of any simulation model. As Kimbrough (forthcoming) argues, "[Frequently, simulation] studies are subjected to the criticism that many sets of decision rules can yield equivalent behavioral outcomes; that is, although a given set of rules yields life-

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

My agents operate by applying proven hill-climbing and reciprocity heuristics to discover trade and the benefits of specialization, and these same principles guide their discovery of theft and their decisions about whether to steal or trade. Since these general behavioral motivations are sufficient to characterize human behavior in discovering and implementing trade, I hypothesize that they should also be sufficient to predict human subjects' implementation of theft. Surprisingly, agents employing these heuristics alone fail to predict human behavior in the new imperfect property enforcement environment; the simple model paints too bleak a picture and predicts degeneration of specialization and exchange as theft crowds out cooperation. Instead of abandoning trade, human subjects innovate to construct a novel property arrangement that facilitates exchange and specialization. They exploit the absence of property protection to develop a second method of trading (consensual taking, or "steal trading") reflected in a specific property convention that emerges in five of six sessions. However, when the potential to adopt that convention (also guided by reciprocity) is added to agents' behavioral repertoire, they again replicate human behavior.

Subject behavior highlights the importance of shared beliefs and the creation of conventions to support exchange. The social interpretation of the act of taking goods from another individual depends powerfully on the conventions in which the act is embedded. Thus, what is interpreted as theft in one case is interpreted as one-half of an exchange agreement under a different property convention. The results suggest that in open-ended environments, an accurate decision-model will require the introduction of agents that form beliefs about the beliefs of others and seek to coordinate those beliefs to form conventions that provide consistent interpretation of actions taken. In the case of property, agents would develop beliefs about which takings constitute violations of property, and then their behavioral heuristics would guide them to develop respect for property (or not) under the chosen interpretation.

Section 2 describes the economic environment and the design as well as results from the model developed in Kimbrough (forthcoming). Section 3 details the notion of property as a convention. Section 4 explains how possibilities for theft and the emergence of property are added to the Kimbrough model and compares the new, imperfect property protection model to the original to define hypotheses for a set of new human subject experiments. Section 5 describes the results of the experiments and compares them to the model predictions. Section 6 details a third version of model as updated by observations from the experiments, and Sect. 7 concludes and summarizes the findings.

2 Simulation and experimental environment

2.1 The economic environment

The underlying economic environment in these simulations and experiments is described in detail in CSW, Kimbrough et al. (2010), and Kimbrough, and all the same features are retained here. Subjects (and agents) are assigned one of two types, *odd*

like outcomes, this does not imply that outcomes in the real world are generated by the same (or similar) process." To establish the validity of the model, it must be able to make out-of-sample predictions.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

and *even*. These types define production functions with increasing returns to one of two goods and Leontief preferences over the goods, with a stronger preference for the good in which they possess increasing returns to specialization. In each trading period, subjects choose a rate of specialization that defines the proportion of their time budgets they will spend producing each good. Subjects know the form of their preferences, but are given no information about their production function except what they discover by experimentation.

Goods appear in subjects' "fields" as they are produced and must be moved by pointing and clicking with a mouse into "homes" in order to be consumed. On each home and field is displayed the number of each good therein, so subjects can observe production and consumption behavior of other subjects. CSW fully protect property in all goods in subjects' homes and fields; that is, no one can take goods from another person at any time. Subjects may exchange by moving goods to other subjects' homes and fields, but the instructions do not inform them of this possibility. Over 35 trading periods of 90 seconds each, subjects learn (or not) to specialize and exchange by trial and error and via communication in a shared chat room. Fully specialized *odd* (*even*) subjects can earn 90 (80) cents per period if they specialize completely and trade with a suitably specialized partner of the opposite type. Subjects working in autarky can maximally earn roughly 1/3 of what they can earn by trading. Hence there are strong incentives to implement specialization and exchange if the possibilities are recognized.

2.2 Learning to specialize and trade

The economic environment in CSW was developed to highlight the discovery process by which individuals implement welfare-improving specialization and exchange. The theoretical apparatus of economics frequently assumes away questions about the sources of discovery (i.e. agents already have perfect information, technology of production is given, etc.), and CSW sought to emphasize the fact that while opportunities to gain from exchange are ubiquitous, there is no reason to assume that such possibilities are immediately known or easily inferred. Indeed some individuals and groups are very successful at discovering and implementing trade in their environment, while others lag behind. Furthermore, the rate of success depends crucially on how groups are formed. CSW observe that it is critical to the regular discovery of specialization and trade that each subject find a suitable trading partner. Large groups, formed all at once, are too chaotic for individuals to find suitable trading partners and reap the gains from trade, but when subjects begin the experiments in pairs and are slowly merged into a larger group (the "Build" treatment), these problems are mitigated and large groups are able to develop and sustain specialization and exchange.

In Kimbrough (forthcoming), the author designs simulated agents to replicate human behavior in CSW's baseline, single large group environment, and then asks how well the simulation model predicts behavior in the sessions in which the groups were built from smaller groups. Model details are included in Appendix A in electronic supplementary material. Like the original CSW experiments, the simulations are performed exactly as in the original model; the only change is the manner in which groups form. Figure 1 below compares rates of efficiency in the simulations of Kimbrough to that of the human subjects in CSW. Figure 1(a) displays efficiency

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

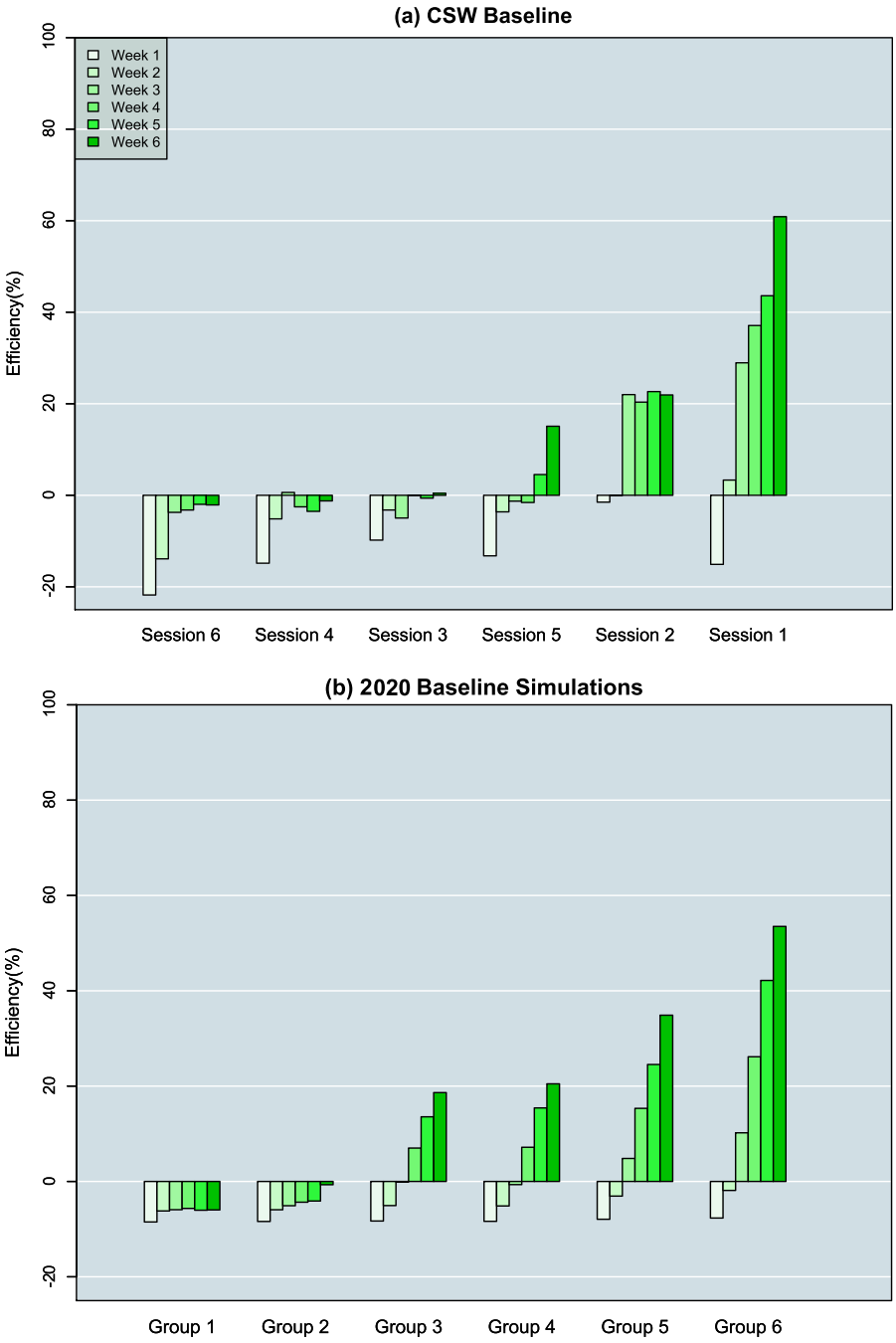


Fig. 1 Average efficiency by session (group) and week—CSW experiments and Kimbrough simulations

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

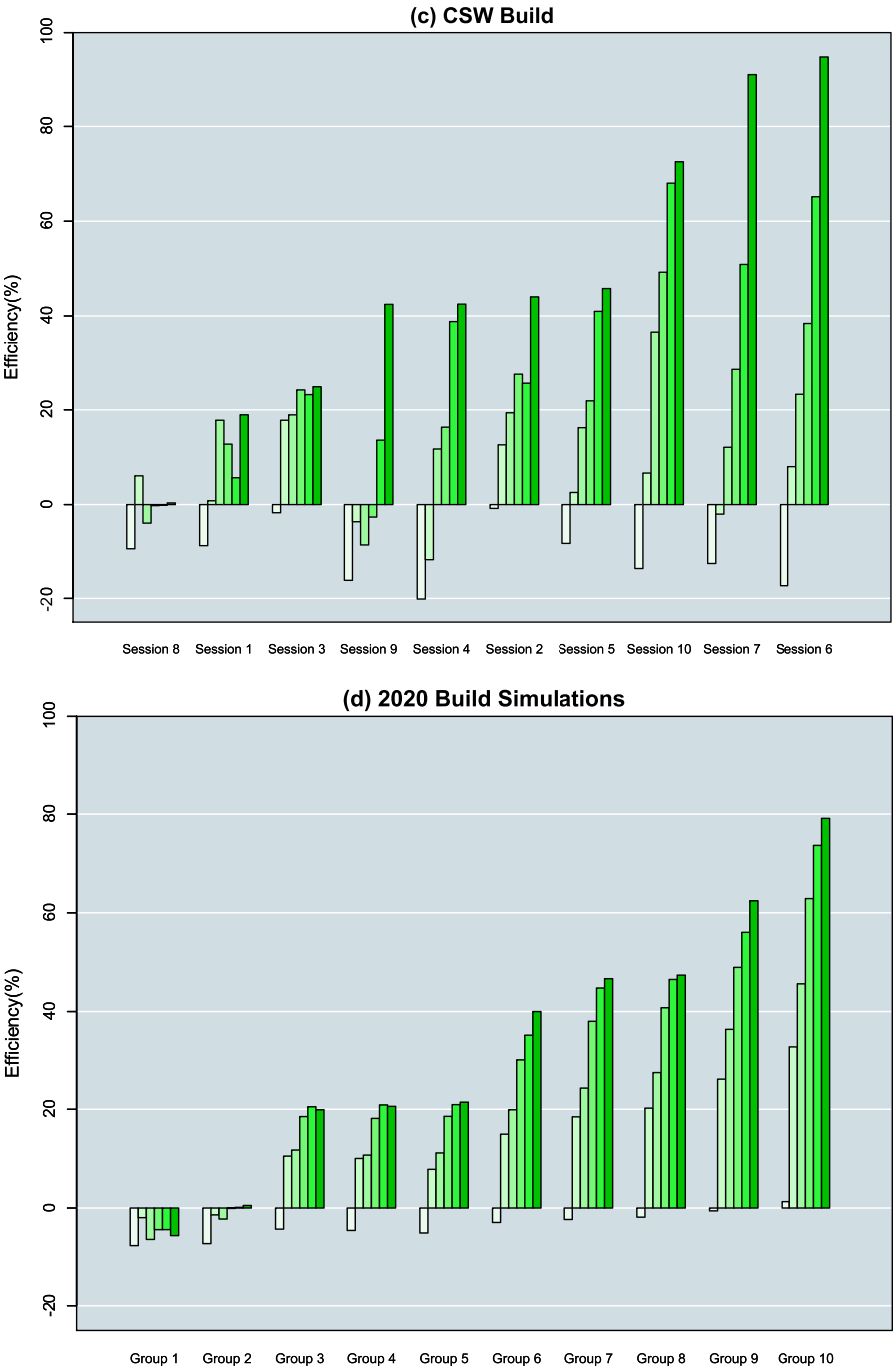


Fig. 1 (*continued*)

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

over the course of each session of CSW's baseline experiments (averaged over six "weeks" and ranked by efficiency), and Fig. 1(b) displays the average simulation efficiency for 6 blocks of 300 simulation runs of one characteristic parameterization of the model, grouped by final period efficiency.[3] As reported in Kimbrough, 95% confidence intervals of mean efficiency from the simulation output contain the mean of human subject data, and 100,000 Wilcoxon tests comparing random samples of 6 simulated sessions to the 6 experimental sessions reject the null hypothesis of equal mean efficiency in more than 5% of tests only in early periods. Figure 1(c) displays efficiency data from CSW's 10 "Build" sessions, and Fig. 1(d) displays the same data from a single parameterization of Kimbrough's Build simulations, displayed in 10 blocks of 180 simulations, again ordered by final period efficiency. For a slightly restricted parameter space, the simulation model still replicates CSW's observations.

The simulations account for human behavior not in terms of the equilibrium properties of the system as a whole but rather via heuristic decision rules based on observed behavior in the experimental environment. While simulated agents are unable to communicate via chat room like the human subjects (I have not, nor could I have, endowed them with the ability), the simulations abstract from the social aspect of chat room exchange, instead modeling the reciprocity heuristics underlying such exchanges. These heuristics identify the *process* by which exchanges (economic and social) are transmitted into behavior in the experimental sessions. As the author argues:

> Basing a model on a controlled experiment (...) allows precise calibration to observable human behavior. In designing experiments, economists induce preferences (Smith 1974, 1982) and production functions, and behavior is bounded and directly recorded. Clear benchmarks exist for optimal behavior, and the process by which subjects achieve, or fail to achieve, optimality is readily observed. The additional detail in data afforded by experiments means that observed data can be readily mapped to agent-level decision rules, and because additional controlled experiments can be performed, the agent-based model can be verified by using it to make predictions about a new experimental environment. If the agents' decision rules yield model output that is accurate in its depiction of human behavior in one environment, the model should be a good predictor of human behavior in other environments. (Kimbrough forthcoming)

The agents employ simple hill-climbing and reinforcement learning heuristics to grope over the production space and develop trading relationships, and the predictive (or replicative) power of these heuristics is robust to an exogenous change in the way groups form in this environment. The next step, then, is to ask whether these agents and the heuristic principles driving their behavior can account for the emergence of other phenomena, namely respect for property.

[3]The parameters chosen for this figure are representative of output from a broad range of parameter choices. See Kimbrough (forthcoming).

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

3 Property as a convention

As Hume (2000) describes the origins of property, "It is only a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules. I observe, that it will be for my interest to leave another in the possession of his goods, *provided* he will act in the same manner with regard to me" (1740, pp. 314–315). Many societies have developed complex mechanisms to punish violators of such rules, and in general, students of property have focused on the creation of enforcement mechanisms when attempting to account for property's origin.[4] However, an account of enforcement institutions is not an account of the emergence of property, for any use of enforcement implies that property has already been violated. Punishment may satisfy a need for retribution (Levine 1998) or incentivize future adherence to a norm or convention (Fehr and Gachter 2000; Bernhard et al. 2006), but both of those hypotheses about the purpose of punishment imply the prior existence of some rule, the violation of which merits punishment. The question remains how to account for the emergence of this original rule and to explain how that emergence can be observed in experimental and agent-based environments. When we consider that the original rule is of the form of a convention as described by Hume, the explanation becomes clearer. In the case of property, the convention may consist of deciding to trade rather than to take in order to augment one's consumption, but this will only be successful *provided* other agents have adopted the same rule.

Young (1993, 1996) details the game-theoretic emergence of such conventions (by his definition, convergence to a single equilibrium when many exist) among boundedly-rational agents for a broad class of games. Repeated interactions may (by chance) give an edge to a particular strategy creating a positive feedback loop whereby agents employing that strategy are more successful than others, and the more agents employ the strategy, the more successful it becomes. In relatively simple agent-based models of games such as the prisoner's dilemma and the stag hunt, it has been demonstrated that naïve agents can converge on simple conventions that come to pervade a population through replicator dynamics (virtual natural selection) or by the epidemiological transmission of behavioral rules (Axelrod 1997; Skyrms 2004). Here I seek to explore the emergence of property conventions in a complex specialization and trade environment, and it will be important to have criteria by which to judge whether a convention has emerged.

Kimbrough et al. (2010, hereafter KSW) explore CSW's Build environment in the absence of exogenous property protection. They alter the CSW environment by eliminating exogenous restrictions on who may view and click on the content of subjects' homes and fields, and they compare various costly enforcement mechanisms to a baseline in which only moral suasion may be used to enforce one's claims. Surprisingly, they find little evidence (1) that a lack of exogenous enforcement reduces efficiency (because some sessions are cooperative without it) and (2) that their mechanisms make subjects better off. In fact, the mechanisms are far

[4]Ranging from Bentham (1802) and Westermarck (1908) to Wyman (2005) and Levine (2005) accounts of property's origin have focused on the *legal* aspects of property and the explicit recognition of rights.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

more likely to do harm through costly retaliation. KSW argue that the relative efficiency of these regimes results from socially created, informal property conventions.

Following Hume (2000), who wrote that "property is nothing but a stable possession, derived from the rules of justice, or the conventions of men" (1740, pp. 324–325), KSW define property empirically as an agreement or convention (either implicit or explicit) that creates stable possession in their experimental environment. They argue that although property, rooted in convention and observable only in its effects, may not exhibit the traits commonly associated with property rights in the folk wisdom (e.g. enforceable contracts, explicit punishment mechanisms, or arbitration), the fundamental fact of property is that it implies the absence of unwanted appropriations of goods. Property, like money, is a self-referential social *practice* – it can only be explained by reference to itself.[5] Thus, *property emerges empirically as the absence of undesired unilateral takings*.[6] This is crucial because this definition of property permits observing the endogenous emergence of an institution in both experiments and agent-based models. With agents, as with human subjects, to understand how respect for property emerges, one must understand what behavioral rules and what sorts of interactions lead to a cessation of takings.

Kimbrough's agents employ simple heuristic learning methods derived from direct observation of the social and economic interaction of human subjects and are able to replicate and predict subjects' propensity to discover and exploit specialization and trade. Agents employing rules of reinforcement learning reciprocity in exchange and trial-and-error, hill-climbing in specialization capture the variety of human behaviors in CSW's experimental environment. My simulation builds upon this research to model how agents develop (or fail to develop) property conventions to support exchange. I hypothesize that a model that extends these heuristics to the process by which agents choose to either take or trade will predict human-subject behavior in attempting to solve the same social problem.[7]

While KSW have previously performed other human-subject experiments with imperfect property enforcement, the Build environment they employed was designed specifically to increase the social cohesion of the group—hence, perhaps, the relative success of subjects in their environment. Here, I will ask how agents perform when they interact in groups of eight from the outset, creating an opportunity to test the model against a new set of experiments.

[5] Bloor (2002) explains this as follows: a metal disc is a coin only in the context of money. Without a conventional notion of what it means for a disc to be a coin—embodied in its use to complete transactions—a metal disc would not be money. It would merely be a metal disc. In the same vein, what is property if others freely expropriate it? Howitt and Clower (2000) develop an agent-based model that yields an emergent market economy with universal adoption of commodity money.

[6] As a referee points out, to use "theft" here instead of "undesired unilateral takings" would mean to assume the existence of the institution this paper seeks to explain. A unilateral taking is only a "theft" if interpreted as such, and this point will become all the clearer in the results below.

[7] The ability to make out-of-sample predictions is the validation criteria for any model.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

4 Extending the model and simulation results

4.1 Implementing theft

The underlying simulation model in this no property protection environment is exactly the same as that described in Kimbrough (forthcoming). Agents explore their economic environment incrementally, learning specialization and exchange via hill-climbing and reinforcement-learning (reciprocity). The sole change is the addition of a function that permits agents to take goods from other agents unilaterally and without engaging in trade. Hereafter, such takings are referred to as "theft", and the new model will be referred to as the T-model. The following pseudocode gives a general overview of the model, and the italicized lines indicate additions to the Kimbrough model:

Model Pseudocode:

```
Set Global Parameters
Initialize Agents
Begin Loop Over Periods
    Begin Period
        Loop 1—Production and Trading Partner Selection
        Loop 2—Theft and Update Theft Probabilities
        Loop 3—Autarkic Consumption
        Loop 4—Trade and Update Trade Probability
        Loop 5—Update Learning, Specialization, and Willingness to Trade
    End Period
    Record Data
End Loop Over Periods
```

To introduce theft, it must be determined, first, whether each agent will steal; second, from whom agents that choose to engage in theft will steal; third, what impact this has on agents who are stolen from; and fourth, what rules might allow some sets of agents to overcome theft and develop respect for property. In general, the initial probability of theft should be non-zero and in some manner based on empirical data; being stolen from should incite retaliation; and, following Hume's notion that conventions emerge *conditionally*, agents should have some chance of crowding out theft via mutually beneficial trade.

Keeping those considerations in mind, theft in the model operates as follows. Agents are initialized either with or without a propensity to steal. Then, with some probability each agent with a propensity to theft will *actually* steal and will acquire all the goods produced by their target agent. Agents who are stolen from, even those with no initial propensity to steal, will become more likely to steal in the future and will direct their future thieving efforts at those who have stolen from them. On the other hand, to offset and potentially crowd out theft, trade relationships will diminish the future probabilities of theft and heal the cracks in inter-agent relationships. Theft occurs after production, but before consumption and exchange. The operational details of the extended model follow.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

4.1.1 Deciding who steals

As stated above, it is important that a randomly-instantiated agent's initial probability of theft be based in some way on empirical data. Some subjects in KSW's experimental treatments begin to steal goods from others almost immediately; others only begin to steal once they've been stolen from; and some never steal at all. As a heuristic with which to construct the agents, I employ the likelihood that an experimental subject engages in theft before using the chat room in an attempt to communicate. This seems a reasonable choice for two reasons.

First, given that their subjects are explicitly made aware of the chat room and must actually experiment with the interface to learn that theft is possible, such behavior suggests something about a subject's approach to the task. Second, subjects from KSW who talk before stealing earn, on average, \$0.11 (roughly 33%) more per period than those who steal before talking, suggesting that the heuristic is useful for categorizing experimental subjects. Thus, because 118 of the 192 subjects in the various treatments of KSW engage in theft before they attempt to communicate, each agent i is instantiated with a variable $thief_i\{0, 0.3\}$ with $P_i(thief = 0.3) = 118/192$, meaning that roughly 60% of agents will attempt to steal in the first period of the simulation. I choose 0.3 because it is roughly equal to one divided by the average period in which KSW's experimental subjects first engage in theft.[8] Thus, of the 60% of agents that attempt to steal, on average 30% will actually steal in the first period of a simulation.

After instantiation, any time an agent is targeted for theft, $thief_i$ is incremented by a value called *stealProbabilityIncrement* to increase the probability of future theft.[9] Furthermore, if an agent has $thief_i = 0$, being stolen from increments this variable and adds another potential thief to the population. On the other hand, any time an agent i engages in trade with another agent j, both $thief_i$ and $thief_j$ are decremented by $2 \times$ *stealProbabilityIncrement* to reduce the probability of future theft. The idea is that agents will engage in both positive and negative reciprocity, but that agents are more sensitive to the opportunity to forgive past wrongs for the prospect of future benefits.[10] Thus, it is possible that all agents will see their probabilities of theft fall to 0 if enough trade occurs.

4.1.2 Rules to determine a thief's target

As in the case of trade, each agent i stores a discrete probability distribution $S_{i,j}$ specifying the likelihood of agent i stealing from each other agent j in I. In each period, those agents for which $thief > 0$ make a random draw, z, from a $\text{uniform}_{[0,1]}$

[8] It is clear that an increase in this value will lead to an increase in theft and a diminishing of cooperation, so the initial value of this parameter is not systematically varied in the simulations described below.

[9] I fix the value of *stealProbabilityIncrement* at 0.1 for all reported T-model simulations.

[10] The double impact of cooperation is partly a practical attempt to "give peace a chance" because theft is much easier than trade in the model. It only takes one person to steal, but for a trade to occur two agents must each select the other as a trading partner. The notion that people tend to be forgiving for the prospect of gains is based on subjective observation of human behavior in the CSW and KSW experiments, but the rule could be adjusted to examine its impact.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

distribution and compare it to *thief*. If $z_i < thief_i$, agent i will choose a target j from the set of other agents with probability $S_{i,j}$. Thus, each agent j will be chosen as the target of theft by agent i with probability $= thief_i \times S_{i,j}$. In the first period, for any agent with $thief > 0$, $S_{i,A} = S_{i,B} = \cdots = S_{i,j,}$ and $S_{i,i} = 0$ for each agent, and for any agent with $thief = 0$, $S_{i,A} = S_{i,B} = \cdots = S_{i,j,} = S_{i,i} = 0$. These probabilities are altered over the course of the simulation by the following process: (1) if agent B steals from agent A, $S_{A,B}$ is incremented by *stealProbabilityIncrement*, augmenting the probability that A returns the favor and steals from agent B in the future; and (2) if agent A trades with Agent C, $S_{A,C}$ is decremented by $2 \times stealProbabilityIncrement$.[11] Note the similarity of these effects to that on the probability of theft in general.

4.1.3 What happens to stolen goods?

Recall that all theft decisions occur prior to the consumption and exchange portions of the model. Once an agent elects to steal from another agent, that agent takes all of the target agent's goods and treats them as his own for consumption and trade purposes. In the *experimental* environment, subjects may steal from as many other subjects as they like, but they are potentially limited in their effectiveness by the prospect of real-time retaliation. Because in the agent-based version of the environment presented here theft must happen sequentially and not in real-time, and because theft may be cumulative (i.e. if A steals from B and then C steals from A, C acquires both A's and B's goods), I randomize the order of theft in each period. Furthermore, I allow agents to steal from *only one* other agent.[12] Thus, it is possible (if all agents are stealing, and the ordering of theft is just right) that a single agent will end a period with all of the goods produced by all agents in that period. It is also possible that an agent will attempt to steal from an agent whose goods have already been stolen. This will not contribute to the breakdown of cooperation (that is, it will not adjust any of the relevant probabilities) because no actual goods change hands.

4.2 Simulation results and experimental hypotheses

I employ the six parameterizations from Kimbrough under the new T-model, and the next section reports results on 1800 simulations of 35 periods each under each parameterization and compares these to the original *No T*-model. I compare the models in terms of efficiency and specialization, and the data from the T-model form my hypotheses for the new experiments. Let π_{it} denote the realized earnings of agent i in period t and π_i^a and π_i^c denote, respectively, expected earnings in autarky and at the global optimum for agent i. Define efficiency in period p for the agents of group N as $\frac{\sum_{i \in N} \pi_{it} - \sum_{i \in N} \pi_i^a}{\sum_{i \in N} \pi_i^c - \sum_{i \in N} \pi_i^a} \times 100\%$. And if q_{it} denotes the total amount of goods produced by

[11] $S_{i,j}$ is bounded below by 0 due to logical constraints. I assume that agents have perfect memories, i.e. that these probabilities do not fade over time.

[12] While this process does not model the human subject environment precisely, it captures the essential features of theft in KSW's experiments in that theft creates a zero-sum environment for the involved parties.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Table 1 Theft (T) treatments—parameters, efficiency and specialization

Treatment	MinEarnAlone	MinEarnTrade	Efficiency	Specialization
1015T	10	15	−11.44%	37.22%
1515T	15	15	−10.80%	37.48%
2015T	20	15	−10.36%	37.47%
1020T	10	20	−10.82%	37.22%
1520T	15	20	−10.34%	37.50%
2020T	20	20	−10.19%	37.53%
1015	*10*	*15*	*2.83%*	*44.23%*
1515	*15*	*15*	*4.38%*	*42.93%*
2015	*20*	*15*	*4.28%*	*41.66%*
1020	*10*	*20*	*3.21%*	*43.72%*
1520	*15*	*20*	*3.53%*	*43.76%*
2020	*20*	*20*	*4.58%*	*41.99%*
CSW	**NA**	**NA**	**4.24%**	**44.49%**
Theft	**NA**	**NA**	**0.05%**	**50.24%**

Italicized entries are from Kimbrough (forthcoming)

Bolded entries are human subject experiments

agent i in period t, and $\bar{q}_i$ is the maximum amount of goods that agent i can produce when fully specialized, define the rate of specialization as $\frac{\sum_{i \in N} q_{it}}{\sum_{i \in N} \bar{q}_i} \times 100\%$.

Table 1 below lists each of the six parameterizations of the T- and *No* T-models and their average rates of efficiency and specialization as well as average efficiency and specialization in the original eight-subject CSW experiments and the new *Theft* experiments reported here. As expected, the introduction of theft to the model has a strong negative impact on both efficiency and specialization. Average efficiency is -10.66% in the T-model and 3.80% in the *No* T-model, and average rates of specialization are 37.40% and 43.05%, respectively. Mann-Whitney tests confirm that the differences are statistically significant.

Finding 1 The T-model simulations are significantly less efficient than the no-T simulations in weeks 3 and 6.

Following CSW and KSW, each 35 trading-period simulation (and experiment) is divided into 6 weeks. Week 3 efficiency measures the performance of each session at the halfway point, and week 6 efficiency measures output by the end. Table 2 reports two-sided Mann Whitney Tests comparing each T-model parameterization to its corresponding no-T model in both week 3 and week 6. These tests reject the null hypothesis of equal mean efficiency for every parameterization; and the difference in means suggests that efficiency is lower in the T-model. Additional, unreported Mann-Whitney tests reject the null hypotheses of equal rates of specialization in week 3 or week 6 for all of the treatments. Goods are also produced at a significantly lower rate with theft. As a robustness check, I also compare the T-model to the CSW

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Table 2 Two-sided Mann-Whitney results comparing efficiency of the theft (T) and no-theft (no-T) simulations by treatment (1800 each)

Treatment	Week 3 (efficiency)	Week 6 (efficiency)
1015T	$U = 2689740^{*}$	$U = 2906682^{*}$
1515T	$U = 2880796^{*}$	$U = 3034669^{*}$
2015T	$U = 2976736^{*}$	$U = 3069754^{*}$
1020T	$U = 2920274^{*}$	$U = 2990344^{*}$
1520T	$U = 2897438^{*}$	$U = 2979198^{*}$
2020T	$U = 3017562^{*}$	$U = 3073506^{*}$

*Significant at $= 0.001$

Table 3 Strength of respect for property T-model (week 6)—number of sessions with perfect, strong, and weak property

Treatment	1015T	1020T	1515T	1520T	2015T	2020T
Perfect (0%)	9	6	7	6	9	8
	(0.5%)	(0.33%)	(0.39%)	(0.33%)	(0.5%)	(0.44%)
Strong (<10%)	50	44	38	37	25	27
	(2.78%)	(2.44%)	(2.11%)	(2.06%)	(1.39%)	(1.5%)
Weak (<20%)	42	48	30	45	23	26
	(2.33%)	(2.67%)	(1.67%)	(2.5%)	(1.28%)	(1.44%)
Total	**101**	**98**	**75**	**88**	**57**	**61**
	(5.61%)	**(5.44%)**	**(4.17%)**	**(4.89%)**	**(3.17%)**	**(3.39%)**

experiments, under the hypothesis that the simulations will also be less efficient than the experiments.

Understanding property as an emergent convention revealed by the absence of theft, I define the strength of respect for property in a given period by the percentage of agents that engage in theft. Then, I define three levels of property enforcement over the final week of each simulation run. Simulation runs have perfect property when no agent steals from another agent in the final week, strong property when less than 10% of agents steal on average each day, and weak property when less than 20% of agents steal on average each day.

Finding 2 By week 6, property is respected in only roughly 5% the simulations.

Table 3 reports the number of sessions (out of 1800) that display perfect, strong, and weak respect for property in week 6 of each T-model parameterization. In only two parameterizations does respect for property (the sum of perfect, strong and weak runs) exceed 5%, with an average of 4.45%. The absence of exogenous protection of property leads agents to steal from one another uncontrollably as theft begets retaliation. The failure of reciprocity to create strong property conventions yields wasteful inefficiency and diminishes the returns to specialization.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

I argue that the rules by which these computerized agents learn specialization, exchange, and theft constitute behavioral predictions about incremental learning and the positive and negative feedback that result from economic interactions between agents with limited rationality in a two-good production and exchange economy. The decision heuristics were generated via rational reconstruction of observed human behavior and an iterative process aimed at calibrating behavioral parameters so that agents would achieve outcomes commensurate with human behavior in an environment *with* exogenous property protection. By extending this model to a new environment *without* property protection, I have created a set of predictions (under the specified behavioral assumptions) whose validity can be tested by returning to the laboratory. Together, the aforementioned findings form my hypotheses for the new experiments. *The* Theft *experiments will be indistinguishable from the T-model and significantly less efficient than the CSW experiments due to the absence of effective property protection.*

5 Experimental treatment and results

5.1 The theft treatment

The new experimental treatment, which I call *Theft*, retains all preferences, production functions, and instructions from the experiments described in CSW and Kimbrough. Subjects must discover the possibility for exchange and specialization in order to reap gains from trade. The *only* variation is that subjects may also, by exploring the computer interface, engage in theft. Based on the results of the simulations, I hypothesize that the absence of exogenous property enforcement will lead to a significant decrease in efficiency. Theft will hinder cooperation, diminishing specialization and exchange and increasing waste.

Eight subjects were recruited at random from the undergraduate student body of a private university in the United States to participate in each of 6 experimental sessions. They sat at visually isolated computer terminals and read instructions from the computer screen. Subjects received $7 for arriving to the experiment on time and received their earnings in cash privately at the end of each 90-minute session. The average experimental earnings were $9.69, ranging from a low of $0.48 to a high of $23.71. No subject participated more than once, and no subject had prior experience with a similar experimental environment. Instructions are included in an appendix in electronic supplementary material.

5.2 Experimental results

Table 1 displays average efficiency and specialization over all six sessions of the *Theft* treatment, and Fig. 2 displays average efficiency by week and average specialization by day for each session. Note the high level of efficiency of Session 5 and also that only one session shows negative efficiency in week 6. It appears that the absence of exogenous property enforcement in this experimental environment does not have nearly as powerful a negative effect on efficiency as the results of the simulations suggest. In fact, by the end of session 5, it is more efficient than any CSW sessions!

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

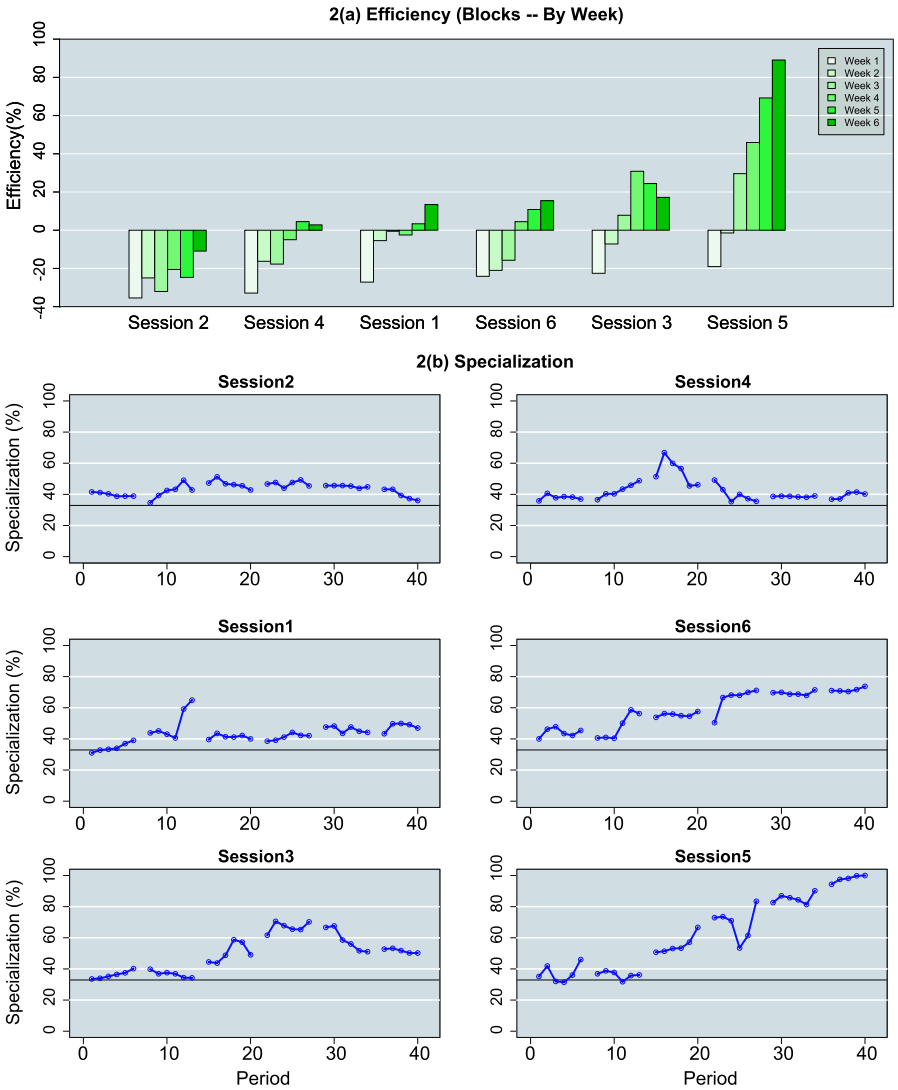


Fig. 2 Average efficiency by week and rate of specialization by day—*Theft* treatment

Finding 3a I cannot reject the null hypothesis of equal mean week 3 and week 6 efficiency for the Theft and CSW sessions.

Finding 3b Nor can I reject the null hypothesis of equal mean rates of specialization in weeks 3 and 6 between the CSW sessions and the Theft sessions.

One-sided Mann-Whitney tests fail to reject the null hypothesis of equal mean efficiency in both week 3 ($U_{6,6} = 23$, p-value $= 0.24$) and week 6 ($U_{6,6} = 16$, p-value $= 0.65$) in favor of the alternative hypothesis that the *Theft* treatment is less efficient

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Table 4 95% bootstrapped efficiency confidence intervals—T-model vs. *Theft* experiments (bolded entries do not contain the *Theft* experimental mean)

	T-model	Theft
Week 1	**[−12.71, −9.37]**	−26.87
Week 2	[−13.54, −7.91]	−12.72
Week 3	**[−13.72, −6.07]**	−4.77
Week 4	**[−14.00, −4.69]**	8.87
Week 5	**[−14.08, −3.70]**	14.61
Week 6	**[−13.96, −2.34]**	21.17

than the *CSW* treatment. Thus, the hypothesis on relative efficiency generated by the T-model can be rejected. Human subjects are able to reap gains from trade, even when property rights are not exogenously enforced.

Furthermore, one-sided Mann-Whitney tests fail to reject the null hypothesis of equal mean specialization in both week 3 ($U_{6,6} = 13$, p-value $= 0.80$) and week 6 ($U_{6,6} = 11$, p-value $= 0.88$) in favor of the alternative hypothesis that the *Theft* means are lower. Thus, contrary to the simulated hypotheses, human subjects also maintain relatively high rates of specialization in the face of insecure respect for property.

By using the T-model to formulate hypotheses about the relationship between the *Theft* and *CSW* treatments, I also create the subsidiary hypothesis that the T-model will be indistinguishable from the *Theft* experiments. However, given that the new experiments show no difference from the original, it is unsurprising that this hypothesis also fails.

Finding 4 The T-model simulations are less efficient than the Theft *experiments*.

Table 4 reports bootstrapped 95% confidence intervals comparing average efficiency in each week over all six parameterizations. The intervals are computed by taking 100,000 samples of 6 T-model simulations each, from each treatment, and then dropping the lowest and highest 2.5% of sample averages to find 95% confidence limits for each week of each treatment. The average of the confidence limits for a given week defines the simulation interval. The experimental mean is contained in the confidence interval only in week 2, and it is actually lower in week 1. However, from week 3 to week 6, the experiments are more efficient than the simulations.

These findings beg the question of why the simulated data of the T-treatment compares so poorly to that of the human subjects in the *Theft* sessions. To answer this question, I return to the experiments and examine subject behavior in more detail. Because I have access to complete data on the flow of goods to and from experimental subjects and their expressed thoughts and intentions (in the form of chat room transcripts), I have a clear view of their behavior as it develops in real-time.

As the experimental sessions unfolded, it became clear that human subjects displayed ingenuity of which the model's agents were constitutionally incapable (they are limited by the foresight of their modeler). Subjects learned quickly that while the ability to take goods from other subjects allowed them to *steal*, it also provided them

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

an additional means of *exchange*. If I take from you with your consent and you take from me with my consent, this is economically equivalent to my actively giving you something in return for something that you give me, and this is precisely the arrangement that a number of the experimental sessions agree upon. In five of the six sessions I observe the emergence a property convention that permits mutual taking *with the same specific content*. The convention is embodied in the following chat transcript segments, each from a different session:

In each of these transcript selections the content of the property convention is clear and specific: subjects agree first to consume the goods they have produced to the best of their abilities by placing them into their homes and *only then* to allow the others to take whatever is leftover in their fields in order to meet their own needs.

Person 7>:	heres what to do
Person 7>:	only take from other people's fields, not houses
Person 7>:	whatever is in the house at the end of the round is what you make money off
Person 7>:	so don't jack other people's house stuff, just fields, and at the end of the round
Person 7>:	feel me?
Person 1>:	yeah, but you can jack from the fields too
Person 7>:	yeah thats what I'm saying
Person 7>:	jack from the fields
Person 7>:	then everyone can still make profit from the houses

Person 6>:	#1, can i steal 9 reds?
Person 5>:	#2 can i take 5 blue?
Person 5>:	#1?
Person 1>:	take from my domino if you need red
(. . .)	
Person 8>:	so now that we've agreed to not steal from each other
Person 7>:	can we take from dominos?
Person 6>:	ok can we take form dominos now?
Person 1>:	yeah, take just from dominos, not houses
Person 5>:	take from other people's dominoes at THE END
Person 6>:	maybe we should have a "grace" period
Person 6>:	like the first 40 seconds don't take from anybody but yourself
Person 6>:	then the last 20 seconds, use the left overs

Person 5>:	the fields are fairgame, lets decide on that from now on, nobody takes ANYTHING from a house, if you have stuff to share you can put it in your field
Person 5>:	your field is your yard sale lol
Person 4>:	lets say after 30 seconds go by though
(. . .)	
Person 4>:	WE SAID THE FIELDS ARE FAIR GAME AFTER LIKE. . . 20 SECONDS

Person 7>:	maybe people should keep their extra on the field and not in their house, and people can take from that?
Person 3>:	that works better for you

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Person 2>:	that's a good idea
Person 4>:	good idea
Person 2>:	then we won't be stealing from each other
Person 7>:	so if you have a whole bunch of red or blue you don't need, move it to the field so people can take it
Person 4>:	i need red. put in field if you don't need it
Person 5>:	i think ppl need to stop pulling from the houses
Person 3>:	k
Person 1>:	lets try not moving stuff
Person 2>:	only add what u need
Person 3>:	add from just the fields?
Person 5>:	ya

Human subjects innovate on their ability to unilaterally take goods from other individuals' homes and fields by adapting the social meaning of "taking" to their circumstances. Rather than treating all takings as malicious violations of property worthy of rebuke and retaliation, subjects come to agree that some takings (particularly those that occur after autarkic consumption has been optimized and explicit trade agreements have been completed) are not actually violations of property at all.[13] By altering the social meaning of property to permit some types of unilateral taking, subjects in the *Theft* treatment are able to achieve unexpectedly high efficiency. On the other hand, my T-model agents are concerned only with the *act* of taking goods or having goods taken from them when making their decisions. The fact that agents in my simulations do not admit the possibility of steal-trading is likely the reason that they do not achieve equivalently high levels of efficiency. With this fact in mind, I develop a third simulation model that permits agents to engage in steal-trading.

6 Steal-trading model

I extend the T-model with a variation called the ST-model (for steal-trading). In addition to stealing from and trading with one another, ST agents may also consensually take unused goods from other agents' homes after the initial consumption and exchange period. Steal-trading agents respect other agents' claims to goods in their "houses", but they consider goods that are leftover in the fields to be "fair game". This additional behavior reflects the innovative convention that emerged in the experiments reported above; by choosing to engage in steal-trading, agents may eliminate wasteful theft and achieve efficiency on par with human subjects.

[13]That this interpretation constitutes a conscious and radical alteration of subjects' initial views on property in this environment is evidenced by the vocal reaction to early unilateral takings in the chat transcripts. For example, by period 4 of Session 5, one subject has already instructed his counterparts to "stop stealing" and "JUST LEAVE EVERYONES STUFF ALONE". In Session 6 a subject laments "everyone just steals it all from everyone else how pointless". But later in both of these sessions, subjects agree that taking is only a problem when goods are moved from houses. Some takings aren't pointless at all.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Table 5 Steal trade (ST) model—efficiency and specialization

Treatment	Efficiency	Specialization
1015ST	4.18%	44.49%
1515ST	5.63%	44.02%
2015ST	4.18%	42.14%
1020ST	1.86%	42.40%
1520ST	4.71%	43.79%
2020ST	5.17%	42.86%

6.1 The mechanics of steal-trading

After agents produce, steal, consume and trade, but before they update their learning rules and record their data for the next period, I introduce the possibility of steal-trading. Only those agents that have a positive probability of theft and are also willing to trade will attempt to steal trade. Since not all agents are willing to steal or trade and since these variables change with experience, I merely provide access to the convention, and there is no guarantee that it will be adopted. Agents that possess unconsumed goods search the set of other agents to determine which other agent has the highest amount of whichever good they are presently wasting the least. The intuition is that an agent steal trades with those agents that can best help it satisfy its preferences, *given* the set of goods it presently possess. If an agent is wasting a larger amount of red than blue, then the agent needs more blue, so it takes from the agent who has the largest amount of leftover blue.

A steal trading agent that has selected a target takes all of the target agent's unconsumed goods and consumes them according to its preferences. If unconsumed goods remain, other steal-trading agents may later take them as part of a second cooperative taking. Furthermore, since steal-trading is simply a second means of exchange, any agent presently unwilling to trade that is party to a steal trade, becomes willing to trade in the next period. Furthermore, if neither thief nor target engaged in trade in the present period, the probability that they trade in the future increases and their probabilities of theft decrease. Thus, both steal-trading *and* trading can ward off theft and crowd out its detrimental effects.

6.2 ST treatments—results

I perform 1800 simulations under the ST-model with each of the six parameterizations used for the T and No-T models. Table 5 displays average efficiency and specialization for each ST parameterization. I group the 1800 simulations for the 1515ST parameterization into six sets of 300 sessions sorted by final period efficiency, and Fig. 3 plots average efficiency by week, and average specialization by day for each of the six groups. Note that in all but the two least efficient groups, efficiency is positive by the end of the session and generally increasing in time.

Furthermore, respect for property is extremely strong under the ST-model. Table 6 shows the number of sessions in each parameterization of the ST-model with perfect, strong, and weak respect for property in week 6. More than 70% of these

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

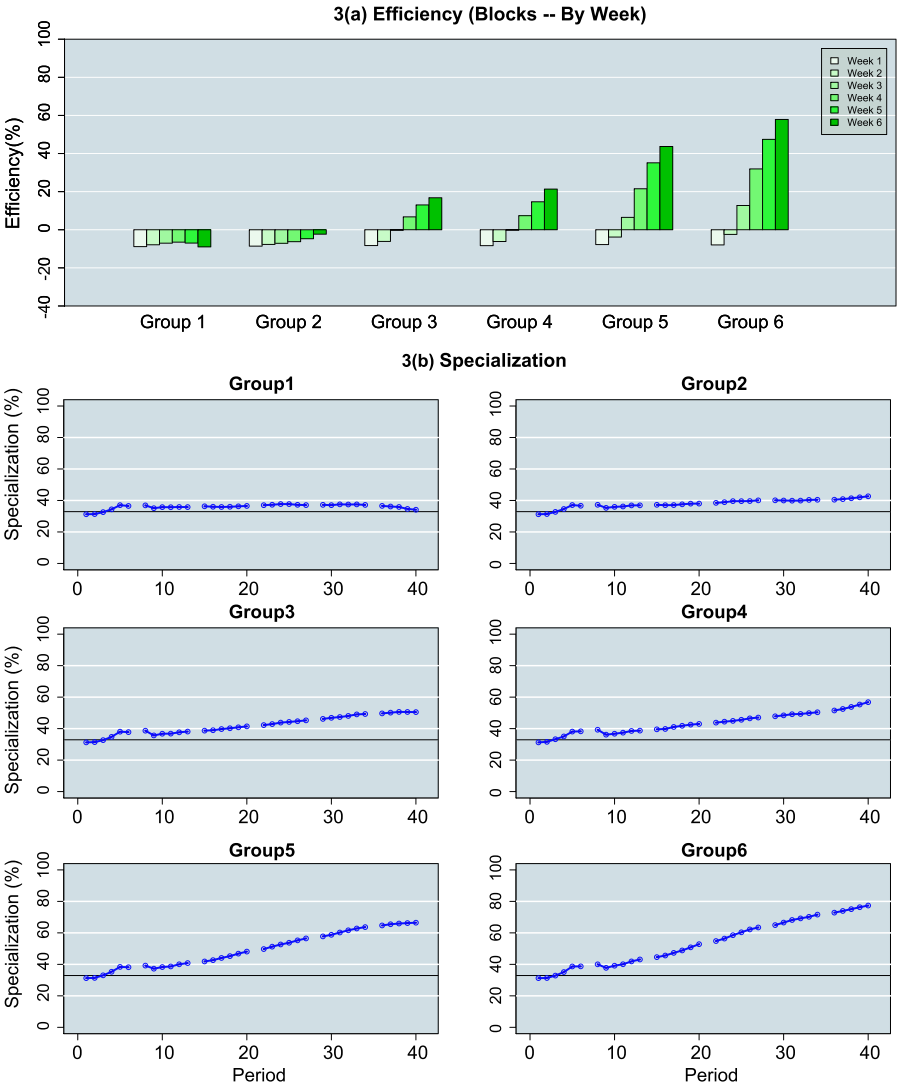


Fig. 3 Average efficiency by week and rate of specialization by day—1515ST treatment

simulations yield perfect respect for property, and on average nearly 95% of sessions respect property at least weakly. The failure of agents to in the T-model to overcome theft in order to reap the gains from trade appears to have been solved by the imposition of the potential for steal-trading. I now compare the ST-model data to the human subject data of the *Theft* treatment. If my observations from the experiments about the importance of steal-trading are accurate, then the new simulation should be indistinguishable from the *Theft* experiments.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Table 6 Strength of respect for property ST-model (week 6)—number of sessions with perfect, strong, and weak property

Treatment	1015ST	1020ST	1515ST	1520ST	2015ST	2020ST
Perfect (0%)	1338	1345	1305	1373	1292	1303
	(74.33%)	(74.72%)	(72.5%)	(76.28%)	(71.78%)	(72.39%)
Strong (<10%)	306	304	318	273	328	323
	(17%)	(16.89%)	(17.67%)	(15.17%)	(18.22%)	(17.94%)
Weak (<20%)	69	64	81	71	74	68
	(3.83%)	(3.56%)	(4.5%)	(3.94%)	(4.11%)	(3.78%)
Total	**1713**	**1713**	**1704**	**1717**	**1694**	**1694**
	(95.17%)	**(95.17%)**	**(94.67%)**	**(95.39%)**	**(94.11%)**	**(94.11%)**

Finding 5 The null hypothesis of equal mean efficiency of the *Theft* experiments and all ST-*model* parameterizations cannot be rejected in weeks 3 and 6.

Table 7 displays bootstrapped 95% confidence intervals for average efficiency in 100,000 random samples of six simulation runs each, for all parameterizations, and compares those intervals to the experimental means from each week of the *Theft* treatment. After week 2, mean efficiency from the *Theft* experiments falls within the confidence interval for *all* parameterizations. Thus it appears that the introduction of steal-trading has solved the problems of the T-model. Furthermore, unreported confidence intervals suggest that, like the *Theft* experiments, the ST-model is indistinguishable from the CSW experiments with exogenous property protection after week 1.

By creating a subset of takings that are not interpreted as theft, agents are able to overcome the temptation of stealing to develop specialization and trade. Thus, it can be argued on the basis of these data that a lack of property protection is insufficient to diminish the productive power of a two-good economy populated with boundedly-rational agents. Because the absence of property permits additional means of exchanging (via steal-trading), subjects and agents are able to eliminate the threat of rampant, costless theft despite their inability to directly protect their goods. If trade and steal trading have some probability of crowding out theft, then high levels of efficiency can attain.

7 Summary and discussion

In Kimbrough agents were created that accurately reproduce the patterns of human behavior in CSWs experimental environment. Here, I extend the agent-based model to a new environment in which the computer program no longer enforces property rights over goods. This model, called the T-model, maintains all features and rules of the original model, but it adds the ability of agents to steal from one another. The T-model yields a sharp decrease in efficiency relative to the original model and also to the original set of experiments. Agents steal from one another and this behavior

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Table 7 95% bootstrapped efficiency confidence intervals—ST model vs. *theft* experiments (bolded entries do not contain the *theft* experimental mean)

Treatment	1015ST	1020ST	1515ST	1520ST	2015ST	2020ST	Theft
Week 1	**[−10.96, −7.04]**	**[−9.76, −6.13]**	**[−9.85, −6.41]**	**[−8.98, −5.77]**	**[−9.04, −5.90]**	**[−8.98, −5.83]**	−26.87
Week 2	**[−11.06, −0.01]**	[−9.92, −0.90]	[−10.12, 1.68]	**[−8.40, 0.72]**	**[−8.92, 0.06]**	**[−8.66, 0.02]**	−12.72
Week 3	[−9.73, 11.59]	[−8.55, 7.63]	[−8.43, 13.11]	[−7.00, 10.15]	[−8.29, 9.24]	[−7.53, 9.80]	−4.77
Week 4	[−6.24, 23.40]	[−7.37, 18.22]	[−4.67, 25.37]	[−4.53, 22.04]	[−5.87, 21.54]	[−4.34, 23.07]	8.87
Week 5	[−2.41, 34.19]	[−4.36, 26.68]	[−0.48, 35.43]	[−1.25, 31.55]	[−1.99, 32.25]	[−0.07, 33.99]	14.61
Week 6	[0.69, 40.93]	[−2.18, 33.73]	[2.39, 41.84]	[1.87, 39.47]	[1.15, 39.59]	[3.07, 41.14]	21.17

escalates uncontrollably—in fact only 5% of simulation runs display even weak respect for property. Because the validity of the model depends on its ability to make predictions about behavior in a novel environment, the results of the T-model constitute predictions for a new set of human experiments employing the same institutional variation.

I perform the second set of experiments (the *Theft* treatment), and another battery of statistical tests rejects the hypotheses produced by the T-model. Not only are the *Theft* sessions more efficient than the T-model simulations, they are also indistinguishable from the human experiments with fully-protected property. Human subjects innovate to exploit a feature of the environment in a way that is impossible for the computerized agents; eliminating exogenous property enforcement permits theft, but it also permits cooperative taking, or steal-trading. The subjects develop a convention that alters the social interpretation of some instances of theft and thereby define a subset of unilateral takings that are permissible. In fact, 5 of 6 human sessions with theft develop a steal-trading convention with the *exact same content*: goods in houses (i.e. goods to be consumed) are inviolable, while goods in fields (i.e. wasted production) are "fair game". Thus imperfect property enforcement allows subjects a second method of welfare-improving exchange.

I then create a third version of the agent-based model (the ST-model) that captures the spirit and effects of this convention. When steal-trading of wasted (i.e. unconsumed and untraded) goods is permitted and such steal-trading may offset the detrimental effects of theft, the ST-model agents once again replicate human experimental behavior. Thus, a third iteration of the model has taken a hypothesis about human social conventions from the observed chat room behavior (which has the potential to be mere cheap talk) and applied it effectively to create agents that mimic human subjects. Statistical tests demonstrate that the model is now indistinguishable from both the human *Theft* treatment and the original *CSW* experiments where theft is impossible. Boundedly-rational agents engage in welfare improving specialization and trade despite the absence of exogenous property enforcement—so long as they employ the appropriate convention. Importantly, when theft and steal-trading are impossible, the steal-trading version of the agent-based model reduces to the original model with perfect protection of property, and it still accurately predicts human behavior in the *CSW* experiments.

Human subject experiments and agent-based models both offer unique views into the processes of economic behavior. Whereas field data on economic systems must be captured in a series of snapshots at various instants throughout a process, both computerized human-subject experiments and simulations allow one to observe the evolution of an economy as it happens. This paper combines the power of the two methods by extracting (or hypothesizing) rules of behavior for simulated agents from observed behavior in the lab.

Whereas Kimbrough derived behavioral rules for individual agents from the actions of lone experimental subjects, this paper extracts an economy-wide property convention to supplement individuals' rules. Respect for property emerges as a convention when subjects make clear the gains from trade, express to others that unilateral takings will only harm their ability to engage in mutually beneficial exchange, and then act in accordance with their stated views. Human groups develop property

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

as part of a process that takes advantage of what Grotius calls our inherent "sociableness" (Buckle 1991). Subjects in the *Theft* treatment employ their sociableness to develop a steal-trading convention that allows them to overcome the incentives for theft and to reap the gains from trade, and it is clear that computerized agents without this social inclination cannot discover such conventions independently. However, when this aspect of human sociableness is imported into the behavioral repertoire of boundedly-rational agents that learn incrementally to specialize, trade, and respect the possessions of others, those agents become indistinguishable from human actors in the same environment.

In that sense, the model is successful because it abstracts from the social aspect of exchange in order to capture the incremental process by which property conventions spread across a population as trade (in the form of both barter and steal-trading) crowds out theft. On the other hand, the model highlights the potential limitations of an agent-based approach to relatively open-ended social problems. Key to the subjects' discovery of steal-trading was their ability to build consensus on the interpretation, or social meaning, of the act of taking goods in various contexts. One way to move closer to an accurate decision model would be to create agents that bargain over and converge to an interpretation of various kinds of takings, but even in such a complex model, the possibility remains that agents would be unable to predict human behavior because the set of possible interpretations must be specified ahead of time. Instead, one should view the method herein as providing a means of testing postulates about what drives the experimental results.

Acknowledgements Thanks to the International Foundation for Research in Experimental Economics, the Mercatus Center at George Mason University, the Economic Science Institute at Chapman University, and the Institute for Humane Studies for generous financial support. Bart Wilson, Taylor Jaworski, Vernon Smith, Daniel Houser, James Gentle, participants at conferences hosted by the Institute for Humane Studies, seminar participants at Maastricht University, Clemson University, and the University of Alabama, and two anonymous referees offered helpful comments and suggestions. Thanks also to Jeffrey Kirchner for his exemplary programming of the experiments and to Jennifer Cunningham for assistance in the laboratory. All errors are my own.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Arifovic, J., & Ledyard, J. O. (2004). Scaling up learning models in public good games. *Journal of Public Economic Theory*, *6*, 203–238.

Arthur, W. B. (1991). Designing economic agents that act like human agents: a behavioral approach to bounded rationality. *The American Economic Review, Papers and Proceedings*, *81*(2), 353–359.

Axelrod, R. (1997). *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton: Princeton University Press.

Bengtsson, H. (2003). The R.oo package—object-oriented programming with references using standard R code. In K. Hornik, F. Leisch, & A. Zeilis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing*. Vienna, Austria.

Bentham, J. (1802). *The theory of legislation*. London: Kegan Paul. C.K. Ogden (Ed.) (1931).

Berkelaar, M. et al. (2008). lpSolve: interface to Lp_solve v. 5.5 to solve linear/integer programs. R package version 5.6.4.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review*, *96*(2), 217–221.
Bloor, D. (2002). *Wittgenstein, rules and institutions*. New York: Routledge.
Buckle, S. (1991). *Natural law and the theory of property: Grotius to Hume*. Oxford: Clarendon.
Crockett, S., Smith, V. L., & Wilson, B. J. (2009). Exchange and specialisation as a discovery process. *Economic Journal*, *119*(539), 1162–1188.
Duffy, J. (2001). Learning to speculate: experiments with artificial and real agents. *Journal of Economic Dynamics and Control*, *25*, 295–319.
Fehr, E., & Gachter, S. (2000). Fairness and retaliation: the economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–181.
Henningsen, A. (2008). micEcon: Microeconomics. R package version 0.5-6.
Howitt, P., & Clower, R. (2000). The emergence of economic organization. *Journal of Economic Behavior and Organization*, *41*, 55–84.
Hume, D. (2000). *A treatise of human nature (1740)*. New York: Oxford University Press.
Kimbrough, E. O. (forthcoming). Heuristic learning and the discovery of specialization and exchange. *Journal of Economic Dynamics and Control*.
Kimbrough, E. O., Smith, V. L., & Wilson, B. J. (2010). Exchange, theft, and the social formation of property. *Journal of Economic Behavior & Organization*, *74*, 206–229.
Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593–622.
Levine, R. (2005). Law, endowments, and property rights. *Journal of Economic Perspectives*, *19*(3), 61–88.
Neuwirth, E. (2007). RColorBrewer: ColorBrewer palettes. R package version 1.0-2.
R Development Core Team (2010). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org.
Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.
Smith, V. L. (1974). Experimental economics: induced value theory. *The American Economic Review*, *66*(2), 274–279.
Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, *72*(5), 923–955.
Warnes, G. R., Bolker, B., & Lumley, T. (2008). gtools: various R programming tools. R package version 2.5.0.
Westermarck, E. (1908). *The origin and development of the moral ideas*. Whitefish: Kessinger (2007).
Wyman, K. M. (2005). From fur to fish: reconsidering the evolution of private property. *New York University Law Review*, *80*, 117–240.
Young, H. P. (1993). The evolution of conventions. *Econometrica*, *61*(1), 57–84.
Young, H. P. (1996). The economics of convention. *Journal of Economic Perspectives*, *10*(2), 105–122.

https://doi.org/10.1007/s10683-010-9258-0 Published online by Cambridge University Press